2062

LUMEA CREATĂ DE INTELIGENȚA ARTIFICIALĂ

TOBY WALSH

Toby Walsh

2062 lumea creată de inteligenta artificială

2062 este anul în care vom avea roboţi la fel de inteligenţi ca noi. Acest lucru este susţinut de majoritatea experţilor în domeniile inteligenţei artificiale şi roboticii. Dar cum va arăta acest viitor? Cum se va desfăşura viaţa pe această planetă?

Profesorul Toby Walsh analizează impactul pe care inteligenţa artificială îl va avea asupra muncii, războiului, politicii, economiei, vieţii cotidiene şi morţii umane.

Pe baza unei înţelegeri profunde a tehnologiei şi a implicaţiilor acesteia, 2062: Lumea creată de inteligenţa artificială descrie alegerile pe care trebuie să le facem astăzi pentru a ne asigura că viitorul va rămâne luminos.

„O invitaţie de neratat la a ne imagina cum vrem să arate viitorul.”

Brian Christian

„Claritate şi raţiune într-o lume ceţoasă şi plină de nesiguranţă – o carte oportună despre cursa pentru a rămâne umani.”

Richard Watson

„Una dintre cele mai profunde întrebări cu care se confruntă umanitatea, cântărită de o minte cu adevărat bine echipată pentru o asemenea întreprindere.”

Adam Spencer

Pentru A şi B,
care îmi fac viaţa atât de plină

0001
Homo Digitalis

Oamenii sunt extraordinari. Într-adevăr, în ciuda vieţii abundente de pe această planetă, probabil că suntem *cea mai* remarcabilă specie care a existat vreodată pe ea. Am inversat cursul râurilor, am construit insule şi am modificat natura după cum am vrut noi. Am ridicat clădiri impresionante. Piramidele monumentale de la Giza. Marele Zid Chinezesc care se

întinde pe o distanţă atât de mare[1]. Superba Sagrada Familia. Am traversat cele mai fierbinţi deşerturi şi am urcat pe cei mai înalţi munţi. Am trimis un automobil în spaţiu. Ba chiar am plecat de pe Pământ şi am mers pe Lună.

Am creat teorii ştiinţifice care explică misterele universului – de la câteva milisecunde de la naşterea sa, ce a avut loc acum 13 miliarde de ani, până la moartea sa inevitabilă, ce se va petrece peste un googol de ani în viitor.[2] Am îmblânzit focul, am eradicat variola şi ne-am unit puterile ca să scăpăm de despoţi şi de dictatori. Am creat artă atât de sublimă, încât oamenii sunt mişcaţi până la lacrimi. Piesa glorioasă a lui Bach *Patimile după Matei.* Frumuseţea golaşă a lui David al lui Michelangelo. Tristeţea tulburătoare a Taj Mahalului.

Dar în ciuda tuturor acestor realizări minunate, în curând vom fi înlocuiţi. Aproape toate urmele de *Homo sapiens* vor dispărea de pe Pământ, la fel cum au dispărut aproape toate urmele predecesorului nostru, *Homo neanderthalensis.* Deoarece evoluţia nu se opreşte niciodată.

---

[1] În ciuda a ceea ce se crede în general, Marele Zid Chinezesc nu este de fapt vizibil din spaţiu. Dar piramidele din Giza pot fi văzute de pe orbita joasă a Pământului, de până la 2000 km altitudine.

[2] Un googol înseamnă 10^{100}, sau 1 urmat de 100 de zerouri. Compania Google a fost numită după o scriere greşită a cuvântului „googol”.

Cu aproximativ 50.000 de ani în urmă, *Homo neanderthalensis* nu a putut sta în calea ascensiunii lui *Homo sapiens.* Nu ştim cu precizie nici când, nici cum au murit oamenii de Neanderthal. Probabil că nu au reuşit să se adapteze la un climat în schimbare – o problemă care ar trebui să sune familiar şi pentru noi în ziua de azi. Sau poate că *Homo sapiens* le-a făcut cu succes competiţie şi nu le-a lăsat nicio nişă ecologică în care să poată supravieţui.

Oricare a fost motivul, este clar că oamenii de Neanderthal au murit şi au fost înlocuiţi de noi. La fel ca oricare altă specie dinaintea noastră, şi noi vom fi înlocuiţi la rândul nostru de una nouă, mai de succes. Şi din moment ce suntem deştepţi – să nu uităm că *sapiens*, sau „înţelept", face parte din numele speciei noastre –, putem chiar să prezicem cine va fi acel succesor.

Succesorul nostru este *Homo digitalis* – specia în care va evolua genul *Homo* în formă digitală. Tot ceea ce facem, precum şi unde facem tot ce facem vor deveni din ce în ce mai mult, şi, în unele cazuri, exclusiv, digitale. Gândurile umane vor lăsa loc gândurilor digitale. Iar activitatea umană din lumea reală va fi înlocuită cu activitatea digitală din lumea artificială şi din cea virtuală. Acesta este viitorul nostru: inteligenţa artificială (IA).

Precedenta mea carte spune povestea inteligenţei artificiale pornind de la grecii antici şi încheind în viitor, la 45 de ani de la momentul scrierii, în anul 2062[3]. Se concentrează pe tehnologie: pe maşinile digitale pe care le construim azi şi care vor începe să fie la fel de deştepte ca noi cândva în jurul anului 2062. Această carte continuă din momentul în care s-a oprit cealaltă[4]. Spune povestea evoluţiei omenirii în secolul sau în cele două secole care urmează după apariţia acestor maşinării gânditoare în jurul anului 2062. Nu se concentrează pe tehnologie, ci pe noi. Examinează impactul pe care maşinăriile care gândesc îl vor avea asupra speciei umane.

Nu vom discuta despre tehnologia ce va fi peste o sută sau două sute ani. După cum ar spune Arthur C. Clarke, tehnologia dintr-un viitor atât de îndepărtat ar

---

[3] În Australia, ultima mea carte a avut titlul *Its Alive! Artificial Intelligence from the Logic Piano to Killer Robots* („Trăieşte! Inteligenţa artificială de la pianul logic până la roboţii ucigaşi"). Dar în Regatul Unit a avut un nume diferit: *Android Dreams: The Past, Present and Future of AI* („Visurile androide: Trecutul, prezentul şi viitorul IA" In Statele Unite s-a numit *Machines That Think: The Future of Artificial Intelligence* („Maşinării care gândesc: Viitorul inteligenţei artificiale").

[4] Iată nişte informaţii interesante despre anul 2062. Elon Musk a prezis că vom avea un oraş format dintr-un milion de locuitori pe Marte până în anul 2062. Întâmplările din *The Jetsons*, îndrăgitele desene animate cu menajere roboţi şi maşini zburătoare, se petreceau la o sută de ani în viitor din anul difuzării lor, în 1962. Pentru cei care au ratat cometa Halley în anul 1986, ar trebui să fie vizibilă din nou privirilor noastre în anul 2062. Dar nu dacă lumea se va sfârşi până atunci: în 1704, Isaac Newton a prezis că apocalipsa va avea loc în anul 2060.

părea pur şi simplu magică[5]. Mai important este *ce facem* cu această tehnologie, pentru că va fi cea mai puternică magie inventată vreodată.

Ascensiunea lui *Homo sapiens*

De ce noi, *Homo sapiens*, am devenit atât de puternici? De ce suntem noi cei care, de bine de rău, dominăm această planetă? De ce *Homo sapiens* l-a înlocuit pe *Homo neanderthalensis*?

Oamenii de Neanderthal nu erau foarte diferiţi de noi. Aveau acelaşi ADN în proporţie de 99,7%. Erau puţin mai scunzi şi mai îndesaţi decât noi, de unde rezultă un raport mai mic dintre suprafaţă şi volum, iar asta înseamnă că erau adaptaţi mai bine la un climat sezonier. Şi în ciuda mitului cu privire la inteligenţa lor, de fapt aveau creiere mai mari decât noi. Dacă iei în considerare diferenţele de mărime, proporţional creierul celor de Neanderthal era comparabil cu al lui *Homo sapiens.*

Atunci de ce am avut noi un avantaj? Probabil că nu vom şti niciodată cu certitudine. Dar un motiv foarte plauzibil este limbajul. Cu aproximativ 100.000 de ani

---

[5] Când eram mic şi visam la construirea maşinăriilor care gândesc, Arthur C. Clarke era unul dintre autorii mei preferaţi de cărţi ştiinţifico-fantastice. Clarke a sugerat că orice tehnologie suficient de avansată nu poate fi deosebită de magie.

în urmă, *Homo sapiens* a început să dezvolte limbajul vorbit complex. Spre deosebire de noi, oamenii de Neanderthal par să fi avut cel mult un protolimbaj simplu. Acest protolimbaj era probabil mai apropiat de muzică decât de vorbire.

Este greu să ne dăm seama dacă aceasta a fost într-adevăr cauza succesului nostru. În cea mai mare parte a secolului XX, nu era considerat un lucru respectabil din punct de vedere ştiinţific să discuţi despre originile limbajului. Având în vedere dovezile limitate ce erau disponibile la vremea respectivă, majoritatea dezbaterilor cu privire la acest subiect erau, în mare parte, speculative. Mulţi considerau că discuţia despre cum s-a format limbajul era de fapt inutilă. Societatea de Lingvistică de la Paris, atunci când a fost fondată în anul 1866, a inclus în legile sale următoarele: „Societatea nu acceptă nicio comunicare în legătură cu originea limbajului sau cu crearea unei limbi universale". Societatea de Filologie din Londra a impus aceeaşi interdicţie în anul 1872.

Aşadar, în cea mai mare parte a secolului XX, lingvişţii au studiat mai mult structura limbajului aşa cum există el în ziua de azi, şi nu s-au gândit prea mult la cum a ajuns să fie astfel. Dar este o întrebare importantă. Cum se face că doar *Homo sapiens* a dezvoltat un limbaj complex? Şi care a fost impactul

asupra evoluţiei noastre?

Chiar şi în anii 70, când începuse să devină din nou acceptabil să discuţi despre originea limbajului, dezbaterea s-a axat mai mult pe felul în care a evoluat acesta, şi nu pe perioada în care s-a întâmplat asta sau pe modul în care ne-a afectat evoluţia. Lingviştii şi-au pus întrebarea dacă limbajul este ceva înnăscut, aşa cum a vrut să ne facă să credem Noam Chomsky, sau dacă a evoluat de-a lungul timpului dintr-un protolimbaj mai simplu. Nu s-a acordat prea multă atenţie impactului gigantic pe care limbajul se pare că l-a avut asupra capacităţii speciei noastre de a domina planeta.

Impactul limbajului

Înainte să existe limbajul vorbit, abilitatea noastră de a învăţa era oarecum limitată. Fiecare generaţie a trebuit să reînveţe foarte multe lucruri de la zero. Desigur, unele cunoştinţe şi unele aptitudini pot fi învăţate prin demonstraţii. Pot să vă arăt plante care sunt otrăvitoare, sau cum să faceţi o suliţă de lemn, sau cum să beţi apă din frunzele din pădurea tropicală. Dar învăţarea prin demonstraţii poate fi o metodă înceată şi dificilă. O persoană trebuie să-i arate fizic alteia tot ce trebuie să ştie. Inevitabil, multe dintre cunoştinţele

unei persoane dispar odată cu ea.

Evoluţia este de asemenea un proces de învăţare, dar unul care este chiar mai încet şi mai puţin precis decât învăţarea prin demonstraţie. Este mai probabil ca genele care duc la comportamente ce măresc rata de supravieţuire a organismului să fie moştenite de generaţia următoare. Dar o astfel de evoluţie are şi ea limitele ei. Vacile nu au evoluat ca să cultive fân. Rechinii nu au evoluat aşa încât să crească foci. Iar fără limbaj este aproape sigur că nu vor reuşi niciodată să facă asta.

Limbajul a schimbat totul. Atunci când a apărut limbajul, o persoană putea *descrie* care plante pot fi mâncate şi care nu: „Nu mânca ciupercile cu pete. Şi nici fructele acestea roşii care arată frumos". Putea descrie şi cum să prinzi o căprioară: „Întotdeauna apropie-te de ea din direcţia din care bate vântul şi cu soarele în spate. La răsărit şi la apus este cel mai bine". Şi cum să cultivi grâu: „Plantează-l primăvara şi culege-l la sfârşitul verii. Aşteaptă să treacă îngheţul, după care-l plantezi din nou".

Dar limbajul nu doar că a ajutat generaţiile următoare să vâneze, să culeagă şi să cultive pământul pentru a se hrăni. Ne-a oferit poveşti, mituri, religie. Limbajul ne-a oferit astronomie şi astrologie, geografie, istorie, economie, politică. Şi ne-a oferit ştiinţă,

tehnologie şi medicină. Limbajul a fost cel care a pus cuvântul *sapiens* în expresia *Homo sapiens.*

Societăţile s-au dezvoltat şi au devenit mai puternice datorită limbii. Aceasta ne-a ajutat să lucrăm împreună, să rezolvăm conflicte şi să avem încredere unii în alţii. Limbajul ne-a permis să dezvoltăm economia prin troc şi, în timp, economia monetară. Limbajul i-a ajutat pe oameni să se specializeze în diverse roluri. A făcut posibil să existe educaţie şi a dus la formarea sistemelor noastre politice.

Cel mai important, limbajul a însemnat că am învăţat nu doar individual, ci şi ca societate. Forţa noastră era acum una *colectivă.* Cunoştinţele nu se mai pierdeau imediat ce murea cineva. Puteau fi comunicate de-a lungul generaţiilor, mai uşor şi mai rapid.

Oamenii de Neanderthal nu aveau nicio şansă în faţa noastră.

Impactul scrierii

Homo sapiens a mai făcut un pas înainte când limbajul a devenit nu doar vorbit, ci şi scris. Acest lucru a reprezentat un al doilea pas ce ne-a ajutat să dominăm planeta.

Scrierea s-a dezvoltat prima dată în China, în jurul anului 5000 î. Hr., şi apoi independent în Mesopotamia, în jurul anului 3100 î. Hr. Oraşele au început să devină

centrele stabile ale vieţii comunitare, iar scrisul fixa legile care le guvernau. Oamenii puteau acum să ţină evidenţa tranzacţiilor sau drepturilor de proprietate şi să creeze un cod penal. Scrierea le permitea oraşelor să funcţioneze într-o manieră mult mai ordonată.

Apariţia scrisului a însemnat că învăţatul nu mai era restricţionat în acelaşi mod de timp sau de spaţiu. Prin limba vorbită puteai să înveţi ceva doar de la cineva care se afla în apropiere, ceea ce te limita la grupul tău social, cu care intrai în contact. Dar odată ce cunoştinţele au putut fi puse pe hârtie, aveai acces la un grup mult mai mare de oameni.

Desigur, iniţial scrierea era un proces încet şi scump. Scribii copiau cu efort textele de mână. Spre exemplu, le lua în jur de o sută de zile să copieze Biblia în acest fel. Majoritatea populaţiei rămânea analfabetă şi profita doar indirect de pe urma beneficiilor pe care le aducea scrisul.

O a treia schimbare a apărut mult mai recent, odată cu tiparul. Johannes Gutenberg a inventat maşina tipografică în jurul anului 1440 d. Hr. În următorii o sută de ani, mai puţin de 100.000 de cărţi au fost tipărite în toată Europa. Dar în secolul ce a urmat, numărul s-a triplat, şi s-a ajuns la peste 300.000 de cărţi. Şi în următorii o sută de ani, producţia de cărţi s-a dublat din nou şi s-a ajuns la aproape 700.000 de cărţi. Azi,

cărţile sunt o industrie de miliarde de dolari, care are sute de mii de angajaţi şi care produce milioane de titluri noi în fiecare an[6].

Cu maşina de tipărit care reducea costul şi timpul necesar producţiei de cărţi, nu a fost nicio coincidenţă că a urmat perioada pe care noi o numim Renaşterea. Ideile şi învăţăturile puteau fi acum răspândite rapid şi uşor. În ziua de azi, internetul ne permite să diseminăm informaţii la nivel mondial cu foarte puţin efort.

Cunoştinţele au devenit ieftine şi abundente. Iar oamenii au devenit mult mai inteligenţi.

Co-învăţare

Următoarea schimbare, al patrulea pas, se întâmplă chiar acum. Este un concept pe care eu l-am numit „co-învăţare". Aceasta este o idee strâns legată – dar uşor diferită – de conceptul de învăţare colectivă.

Sociologii, antropologii şi alţii descriu cum *Homo sapiens* a progresat prin grupuri învăţând împreună de-a lungul generaţiilor. Aceasta este învăţarea colectivă. Fiecare generaţie învaţă colectiv de la cea de dinainte. Ca grup suntem mai inteligenţi, dar niciun

[6] Vom vorbi pe scurt despre schimbările exponenţiale, dar producţia de cărţi a fost una dintre primele schimbări exponenţiale care au avut un impact asupra societăţii noastre, una care a transformat societatea complet.

individ nu e neapărat mai inteligent decât cei din trecut. Co-învăţarea, pe de altă parte, nu reprezintă învăţarea ca grup, ci priveşte *fiecare individ* din cadrul unui grup. În co-învăţare, fiecare persoană învaţă tot ce învaţă şi ceilalţi din grup. Fiecare individ din grup deţine aceleaşi cunoştinţe, aşa că fiecare individ din grup devine mai deştept.

Prin limbajul vorbit, co-învăţarea poate avea loc în cazul unor grupuri de zeci sau sute de oameni. Tu îmi explici ceva, iar eu învăţ. Prin limbajul scris, putem creşte acest număr până la milioane sau chiar miliarde. Tu scrii ceva ce ai învăţat, şi oricine are acces la acea scriere poate învăţa acelaşi lucru. Dar sunt multe aptitudini pe care nu le putem reda cum trebuie altor oameni. Este la fel de greu pentru tine să înveţi să mergi pe bicicletă cum a fost şi pentru mine. Nu pot să spun sau să scriu prea multe ca să fac lucrurile mai uşoare pentru tine.

Şi astfel ajungem la ultima treaptă a schimbării în ce priveşte învăţarea – cea care îi va oferi lui *Homo digitalis* un avantaj de neînvins. Co-învăţarea începe să se desfăşoare nu prin intermediul limbii, ci prin intermediul codului de calculator. Iar acesta este un mod mult mai bun de a co-învăţa: pur şi simplu îţi transmit o copie după codul meu. Nu este nevoie de nicio traducere, nici într-un sens, nici într-altul, iar

codul este executabil direct şi instantaneu. Şi, spre deosebire de memoria noastră, codul nu se pierde. Odată ce a fost învăţat, nu mai poate fi uitat. Este greu să ne imaginăm o metodă mai bună pentru co-învăţare decât transmiterea codului de calculator.

Învăţare la nivel mondial

Companii precum Tesla şi Apple încep deja să co-înveţe la scară globală. Apple, spre exemplu, foloseşte această tehnică a co-învăţării pentru a-şi îmbunătăţi software-ul de recunoaştere a vorbirii. Fiecare telefon mobil inteligent Apple de pe planetă învaţă şi îmbunătăţeşte codul folosit să recunoască vorbirea oricărui alt smartphone Apple. În mod similar, Tesla se foloseşte de co-învăţare ca să-şi îmbunătăţească abilitatea de conducere autonomă. Fiecare maşină Tesla poate să-şi îmbunătăţească felul în care merge şi felul în care merge oricare altă maşină Tesla. În fiecare noapte, maşinile Tesla pot să descarce şi să împărtăşească unele cu altele cele mai recente îmbunătăţiri din software-ul lor. Dacă o Tesla a învăţat cum să evite un cărucior de cumpărături rătăcit, toate maşinile Tesla de pe planetă vor învăţa rapid cum să facă şi ele asta.

Co-învăţarea este un motiv pentru care *Homo sapiens*

nu are nicio şansă în faţa *Homo digitalis.* Este de asemenea şi unul dintre motivele pentru care vom fi surprinşi de viteza cu care va ajunge *Homo digitalis.* Suntem obişnuiţi să învăţăm cam totul de la zero, de unii singuri. Nu avem nicio experienţă personală cu privire la învăţarea la scară planetară.

Imaginează-ţi că ai putea să co-înveţi la fel cum o fac calculatoarele, prin distribuirea codului. Ai putea să vorbeşti orice limbă din lume. Ai putea juca şah la fel de bine ca Garry Kasparov sau Go la fel de bine ca Lee Sedol. Ai putea să demonstrezi teoreme la fel de uşor ca Euler, Gauss sau Erdős. Ai putea scrie poezii care ar rivaliza cu cele ale lui Wordsworth sau ale lui Shakespeare. Ai putea să cânţi la toate instrumentele muzicale. Puse una peste alta, abilităţile tale ar fi asemenea celor mai bune abilităţi ale oricui de pe planetă. Şi ai continua să devii din ce în ce mai bun. Poate că asta sună înfricoşător, dar este viitorul nostru de co-învăţare, ce va avea loc atunci când membrii speciei *Homo digitalis* vor începe să-şi transmită unii altora coduri de computer.

Pentru a înţelege mai bine beneficiile învăţării prin codul-maşină, trebuie să înţelegi încă două idei importante. În primul rând, faptul că toate calculatoarele sunt maşini *universale* care pot să lucreze cu *orice* program. Şi în al doilea rând, faptul că

programele pot să se modifice singure. Mai precis, un program se poate modifica pentru a deveni mai bun la îndeplinirea unei sarcini date. Permite-mi să explic în detaliu de ce aceste două concepte sunt atât de puternice şi de ce îi vor oferi lui *Homo digitalis* un mare avantaj în faţa noastră.

Maşini universale

Alan Turing a fost unul dintre strămoşii IA. A rămas în istorie prin faptul că s-a întrebat ce ar însemna că un calculator ar putea să gândească. El a fost cel care a pus bazele programării. A inventat o idee simplă, dar revoluţionară: maşina computaţională *universală.* Aceasta este maşina care poate calcula tot ce poate fi calculat. Da, ai citit corect. De când Turing a venit cu această idee, am reuşit să construim un calculator care, în principiu, poate calcula tot ce poate calcula *orice* calculator, chiar şi cele care nu au fost inventate încă.

Centrale pentru ideea unei maşini computaţionale universale sunt conceptele *de program* şi de *date* pe baza cărora operează acest program[7]. Programele sunt

[7] Maşinăria universală a lui Turing este un dispozitiv mai degrabă abstract şi oarecum mai mecanic decât calculatoarele pe care le avem în prezent. Cu toate acestea, nu este mai puţin puternic. Constă într-o bandă de hârtie pe care sunt scrise simboluri, un cap care poate să citească banda, să scrie simboluri noi pe bandă şi să mişte banda de la stânga la dreapta, şi o cutie cu electronice care face

succesiuni de instrucţiuni pe care calculatorul le urmează atunci când rezolvă o problemă. Gândeşte-te la ele ca la nişte reţete. Datele sunt informaţiile diferite pe baza cărora acţionează programul – sunt precum ingredientele folosite în timpul unei sesiuni de gătit.

Să ne gândim la problema actualizării soldului unei persoane atunci când face o plată electronică. Putem crea un program care va face acest lucru indiferent de cuantumul plăţii sau de persoana care face plata. Datele după care acţionează programul reprezintă o bază de date privind numele clienţilor şi soldurile lor bancare, plus numele persoanei care face plata şi cuantumul acesteia.

Programul necesar pentru o plată electronică funcţionează astfel: mai întâi, acesta caută numele persoanei şi soldul în baza de date. Apoi scade plata din sold. În ultimul rând, programul actualizează noul sold în baza de date. Este simplu – dar extrem de important. Prin schimbarea datelor, putem scădea plata de la un alt client, sau chiar dintr-o bază de date de clienţi care aparţin unei bănci diferite. Şi dacă schimbi programul, calculatorul poate face ceva nou.

---

diverse acţiuni cum ar fi citirea benzii, scrierea benzii sau mişcarea ei, în funcţie de starea ei internă şi de simbolul citit cel mai recent. Turing a descris pentru prima dată o astfel de maşinărie în anul 1973. Vezi Alan Turing, „On Computable Numbers, with an Application to the *Entscheidungsproblem", Proceedings of the London Mathematical Society*, vol. 42, pp. 230-265.

De exemplu, dacă adăugăm plata în loc să o scădem, avem un program care se ocupă cu depozite electronice, în loc de plăţi electronice.

Atunci, un calculator este un exemplu de maşină *universală* deoarece poate să lucreze cu *orice* program. Poate fi încărcat cu noi aplicaţii, care sunt programe ce îi permit să îndeplinească sarcini pe care creatorii telefonului mobil inteligent nu le-au anticipat. Astfel, smartphone-urile au devenit mult mai mult decât nişte simple telefoane: sunt navigatoare, calendare, alarme, calculatoare, spaţii pentru notiţe, aparate de muzică, console pentru jocuri şi, din ce în ce mai mult, asistente personale.

Progresele în tehnologie ne pot aduce calculatoare mai rapide, dar nu vor putea să calculeze nimic mai mult decât maşinăria universală care a fost visată prima dată de Turing, în anii ’30. Cel mai remarcabil este faptul că Turing a avut această idee a unei maşinării computaţionale universale înainte ca primele calculatoare să fie construite.

Mai mult de atât, calculatoarele sunt *singurele* maşinării universale ce au fost inventate de oameni. Gândeşte-te la ce ar presupune o maşinărie universală de călătorit. Aceasta ne-ar permite să zburăm prin aer, să înotăm pe sub apă şi să călătorim peste tot pe uscat. Ar putea trece peste căi ferate, peste asfalt, peste iarbă,

până şi prin nisipuri mişcătoare. Ar putea transporta o persoană sau douăzeci. Ar putea chiar să zboare până pe Lună. Gândeşte-te la *Transformers*, pe steroizi.

Pentru a îndeplini o sarcină nouă, un calculator are nevoie doar de un program nou. Acest aspect face calculatoarele să fie veşnic adaptabile. Calculatoarele pe care le avem acum sunt capabile să facă mult mai mult decât fac azi. Au până şi potenţialul de a ajunge inteligenţe artificiale. Nu trebuie decât să găsim programul potrivit pentru calculator.

Asta ne aduce la următoarea idee importantă. Nici măcar nu este necesar să *găsim* acel program nou, pentru că poate face asta computerul de unul singur. Poate să înveţe să facă lucruri noi. Poate şi să înveţe să se comporte inteligent.

Maşinării care învaţă

Cum poate un calculator să înveţe ceva nou? La urma urmei, un program de calculator este doar o succesiune de instrucţiuni, specificată de un anumit cod de calculator. De fapt, termenul de *computer code* sau „cod de calculator" este unul bun, întrucât instrucţiunile unui program chiar sunt specificate de un cod criptic. Pe un calculator Z80, spre exemplu, codul 87 înseamnă că adăugăm două numere împreună, în timp ce 78

înseamnă că programul s-a terminat. Iar pe un calculator 6800, codul 8B adaugă numere, în timp ce DD încheie programul[8].

Lucrul important în cazul codului nu este că este criptic, ci că reprezintă date. O succesiune de numere. Aceasta este o idee extrem de puternică. Dacă vrem să schimbăm programul, pur şi simplu încărcăm un nou cod drept date. Şi ce e şi mai important: deoarece un program poate să-şi schimbe propriile date, un program poate *să se schimbe singur.* Aici este miezul învăţării automate: ideea că un calculator poate să înveţe din propriile date şi să-şi schimbe propriul cod pentru a-şi îmbunătăţi rezultatele în timp.

Nu e atât de important să ştii cum algoritmii învăţării automate decid ce schimbări trebuie să facă în propriul cod. Unii sunt inspiraţi de evoluţie, creând mutaţii şi amestecuri ale codului la fel ca mutaţiile şi amestecurile de gene din reproducţia sexuală. Alţii sunt inspiraţi de creierul însuşi, actualizând legături între neuronii

---

[8] Instrucţiunile pentru microprocesoarele Z80 şi 6800 sunt date în numere hexadecimale, sau şaisprezecimi. Numerele decimale sunt zecimi: odată ce treci peste 9, ajungi la 10, 11, 12 şi aşa mai departe. În şaisprezecimi, când numeri peste 9, foloseşti A (=10), B (=11), C (=12), D (=13), E (=14), F (=15), iar apoi 10 (=16), 11 (=17), 12 (=18) şi aşa mai departe. Instrucţiunile DD pentru microprocesorul 6800 sunt cunoscute în mod hilar sub numele mnemonic de „Opreşte-l şi ia foc". Înainte ca astfel de microprocesoare să existe, calculatoarele erau imense şi nu te puteai baza pe ele, şi exista întotdeauna un mic risc ca oprirea unui calculator să îl facă să ia foc.

artificiali, precum întărirea conexiunilor nervoase care au loc în creierul nostru atunci când învăţăm. În ambele cazuri, calculatorul păstrează schimbările care îmbunătăţesc performanţa şi le elimină pe cele care nu ajută cu nimic. Încet, dar sigur, calculatorul învaţă cum să funcţioneze mai bine.

Deja avem un exemplu foarte bun despre cum se poate construi inteligenţa. Suntem noi – *Homo sapiens.* Inteligenţa noastră este în mare parte învăţată. Ne naştem fără să cunoaştem o limbă. Fără capacitatea de a citi şi de a scrie. Fără cunoştinţe legate de aritmetică, sau astronomie, sau istorie antică. Dar putem învăţa toate acestea, şi multe altele.

Se pare că învăţarea automată este o parte importantă a calculatoarelor care gândesc. Este o soluţie pentru *blocajul cunoştinţelor*, problema de a turna într-o maşinărie toate cunoştinţele pe care le-am obţinut de-a lungul a mii de ani. Programarea tuturor acestor cunoştinţe de către cineva, una câte una, ar fi un proces încet şi neplăcut. Dar nu este nevoie să facem asta, deoarece calculatoarele pot învăţa pur şi simplu singure.

Acum înţelegem de ce calculatoarele pot fi mult mai bune decât oamenii la învăţat. Pot crea un program care să ştie cum să-şi îmbunătăţească propriul cod, iar apoi pot folosi acest cod în comun cu mai multe

calculatoare. E simplu! Şi mult mai eficient decât cum învăţăm noi, oamenii.

Data viitoare când vei încerca să-ţi înveţi copilul cum să calculeze maximul unei funcţii matematice sau cum să conjuge un verb în germană, gândeşte-te cât de uşor ţi-ar fi să-l înveţi asta dacă ar fi calculator. Pur şi simplu i-ai da codul necesar.

Calculatoarele fac mai mult decât li se spune

Învăţarea mecanică a alimentat multe dintre recentele progrese spectaculoase în domeniul IA. A făcut ca programul Alphago al Google să poată să-i bată pe cei mai buni jucători umani de Go de pe planetă. Este ingredientul secret din spatele Google Translate. Şi stă la baza multor alte programe care-i pot întrece acum pe oameni la diagnosticarea cancerului de piele şi la jocurile de poker.

O reacţie comună în faţa ideii de învăţare automată este aceea că un calculator va face doar ceea ce este programat să facă. La un nivel simplu, această afirmaţie este corectă. Calculatoarele sunt deterministe[9]

[9] În mod ironic, deşi calculatoarele sunt maşinării deterministe, ştiinţa computerelor este una dintre ştiinţele în care experimentele au cel mai mic grad de repetabilitate. Calculatoarele au devenit nişte sisteme foarte complexe, care funcţionează în reţea. în consecinţă, adesea este imposibil să recreezi cu precizie condiţiile unui experiment anterior.

în totalitate. Urmează instrucţiunile care sunt scrise în codul lor computerizat. Ele nu deviază de la cod – nu pot face asta. Dar la un nivel mai complex, calculatoarele pot face lucruri pe care nu au fost programate în mod explicit să le facă. Pot să înveţe programe noi. Pot să fie chiar şi creative. La fel ca oamenii, pot învăţa lucruri noi din propriile experienţe.

Alphago nu a fost programat ca să joace jocul antic chinezesc mai bine decât un campion mondial. A învăţat cum să facă asta, jucându-l de un milion de ori. Motivul pentru care a devenit mai bun la Go decât oamenii a fost că l-a jucat mai mult decât oricare om în viaţa sa. Şi învăţând cum să joace Go bine, a devenit şi puţin creativ. Foloseşte mişcări la care maeştrii în Go nu se aşteaptă niciodată şi găseşte noi posibilităţi în care jocul poate fi jucat.

Iar Alphago nu este singurul program de acest tip. Calculatoarele sunt mai bune decât oamenii la o mare varietate de jocuri, inclusiv table, poker, jocurile de cuvinte şi şah. Atunci când cineva îmi spune că un calculator poate face doar ceea ce a fost programat să facă, îmi place să enumăr o listă de jocuri la care calculatoarele sunt deja campioane mondiale. În aproape fiecare caz, aceste programe pe calculator au fost programate de jucători cu abilităţi intermediare, şi programul a devenit un campion mondial învăţând

cum să joace mai bine decât oamenii.

Avantajul maşinăriilor

Pentru a înţelege de ce *Homo sapiens* va fi inevitabil înlocuit, trebuie mai întâi să înţelegi numeroasele avantaje pe care un calculator le are în faţa unui om – şi avantajele pe care lumea digitală le are în faţa celei analoge. Co-învăţarea este un avantaj important, dar haide să aruncăm o privire şi asupra altor avantaje.

Primul este acela că un calculator poate să aibă capacitate de memorie mult mai mare faţă de un om. Toate amintirile noastre trebuie să fie stocate în craniile noastre osoase. Într-adevăr, deja plătim un preţ semnificativ pentru că avem nişte capete atât de mari. Până de curând, naşterea era unul dintre motivele principale de deces în rândul femeilor. Şi de fapt mărimea canalului pelvigenital pune limite pentru cât de mari pot deveni craniile umane. Calculatoarele nu au astfel de limite. Putem pur şi simplu să le adăugăm mai mult spaţiu de stocare.

Al doilea avantaj constă în abilitatea computerelor de a lucra la viteze mult mai mari decât oamenii. Creierul funcţionează la sub 100 Hz întrucât neuronii necesită a o suta parte dintr-o secundă pentru a acţiona. Creierele noastre sunt atât chimice, cât şi electrice, ceea ce le

încetineşte şi mai mult. Chimicalelor le ia un anumit timp pentru a trece de graniţele nervilor, şi pentru a avea anumite reacţii chimice. Calculatoarele, pe de altă parte, depind doar de legile fizicii. Viteza calculatoarelor a crescut de la 5 MHz în anul 1981 (adică abilitatea de a executa cinci instrucţiuni la fiecare milionime de secundă) la în jur de 5 GHz în prezent (capacitatea de a executa cinci instrucţiuni la fiecare miliardime de secundă). Desigur, doar viteza în sine nu este un indicator foarte bun pentru performanţa unui calculator în ziua de azi. Viteza ca atare a calculatorului nu a crescut prea mult în ultima vreme. În schimb, calculatoarele sunt mai rapide acum deoarece pot să facă mai multe lucruri în acelaşi timp. Precum creierul uman, computerele pot să execute câteva instrucţiuni simultan. Cu toate acestea, siliciul are un avantaj fundamental al vitezei în faţa biologiei.

Al treilea avantaj pe care îl au maşinăriile este acela că nu au o sursă de energie limitată, precum oamenii. Creierul nostru foloseşte în jur de 20 de waţi din cei 100 de waţi pe care îi produce corpul unui adult[10]. Avantajul evolutiv al inteligenţei justifică investiţia energiei limitate a corpului nostru în creier, dar nu există altă rezervă de energie pe care s-o investim într-o

[10] Creierul foloseşte cea mai multă energie din toate organele noastre. Prin comparaţie, inima foloseşte mai puţin de 5 waţi.

gândire avansată. Prin comparaţie, un calculator obişnuit poate să folosească 60 de waţi de energie. Şi dacă îţi doreşti să ai mai multă energie (sau putere de calcul), poţi pur şi simplu să distribui o parte din sarcini în cloud[11]. Cele şapte miliarde de creiere de pe planetă consumă colectiv în jur de 14 gigawaţi de energie. Prin comparaţie, calculatoarele la nivel mondial folosesc deja de zeci de ori mai mult decât această cantitate. De fapt, calculatoarele reprezintă azi 10% din electricitatea folosită la nivel global, sau peste 200 de gigawaţi. Şi acest număr va continua să crească.

Al patrulea avantaj pe care-l au calculatoarele este că oamenii au nevoie de odihnă şi de somn. Calculatoarele pot să funcţioneze non-stop şi nu obosesc niciodată. Aşa cum am văzut mai devreme, Alphago a devenit atât de bun la Go pentru că a jucat mai multe partide decât orice om. Bineînţeles, pentru oameni, somnul are şi alte scopuri în afară de asigurarea odihnei şi de recuperarea puterii. Ne poate ajuta să ne actualizăm amintirile, spre exemplu, şi să abordăm problemele în mod subliminal. Este posibil ca şi calculatoarele să poată beneficia de pe urma unui

[11] Literalmente „nor“, concept modern în domeniul computerelor şi informaticii, care se referă la un ansamblu distribuit de servicii de calcul, aplicaţii, acces la informaţii şi stocare de date, fără ca utilizatorul să aibă nevoie să cunoască amplasarea şi configuraţia fizică a sistemelor care furnizează aceste servicii. (n. red.).

astfel de proces? Am putea să le programăm să doarmă o zi întreagă din când în când.

Al cincilea avantaj pe care îl au calculatoarele este că ele nu uită nimic, spre deosebire de oameni. Gândeşte-te de câte ori ai pierdut timp căutând un obiect. Sau de câte ori ai uitat o zi de naştere. Fără îndoială, uitatul poate fi uneori un lucru util deoarece ne ajută să ignorăm detaliile irelevante – dar este o problemă banală să programezi calculatoarele să uite.

Al şaselea avantaj este că nu sunt orbite de propriile emoţii, precum oamenii. Computerele de azi nu au emoţii care să le întunece judecata. Pe de altă parte, emoţiile joacă un rol important în vieţile noastre, şi adesea influenţează într-un mod pozitiv felul în care luăm decizii. Aşadar, este probabil să aibă valoare evolutivă. Am putea alege să le oferim calculatoarelor emoţii pe viitor. Voi discuta acest aspect mai în detaliu în capitolul al treilea, împreună cu alte subiecte dificile, cum ar fi conştiinţa.

Al şaptelea avantaj este unul pe care l-am explorat deja: oamenii au limite în felul în care îşi pot pune în comun cunoştinţele şi aptitudinile, dar calculatoarele nu. Orice computer poate să ruleze codul unui alt computer. Când un calculator învaţă cum să traducă din mandarină în engleză, putem să oferim această abilitate tuturor celorlalte. Când un calculator învaţă

cum să diagnosticheze melanomul, această aptitudine poate fi transmisă tuturor calculatoarelor. Ele sunt co-studenţii supremi.

Al optulea avantaj derivă din faptul că oamenii nu se pricep prea bine să ia decizii. Am evoluat până în stadiul în care ne descurcăm să supravieţuim, dar acest lucru este departe de a fi ideal. Spre exemplu, suntem groaznici când vine vorba să calculăm care sunt probabilităţile exacte. Dacă ne-am pricepe la asta, nu am mai cumpăra niciodată bilete la loterie. Dar putem programa calculatoarele astfel încât să fie optime. Domeniul economiei comportamentale se ocupă cu studiul deciziilor noastre mai puţin decât optime. De exemplu, adesea urmărim nu să ne mărim profiturile, ci să evităm să avem pierderi. Experţii în economie comportamentală o numesc *aversiunea faţă de pierdere.* Mai sunt şi alte exemple de comportament iraţional. Mulţi dintre noi se tem să zboare când ar trebui de fapt să ne fie mai frică de mersul cu maşina până la aeroport. Ştim că trebuie să mai dăm jos câteva kilograme, dar gogoaşa aceea cu gem e mult prea delicioasă.

Desigur, lucrurile nu stau doar într-un singur fel. Calculatoarele nu sunt mai bune decât noi din toate punctele de vedere. Oamenii au câteva avantaje majore în faţa computerelor. Creierele noastre încă sunt mai

complexe decât cele mai mari supercomputere. Învăţăm rapid şi avem o creativitate extraordinară, inteligenţă emoţională şi empatie socială. Dar mă îndoiesc că vom mai avea aceste avantaje în faţa calculatoarelor pe termen lung. Deja avem câteva dovezi care demonstrează faptul că şi calculatoarele pot să fie creative şi să posede inteligenţă emoţională şi empatie. Pe termen lung, *Homo sapiens* nu are şanse prea mari într-o cursă împotriva maşinăriilor.

Succesorul nostru

Atunci cine este acest *Homo digitalis*, această specie şi mai remarcabilă ce urmează să ne înlocuiască?

O specie este definită de ceea ce este şi de locul în care acţionează. În cazul lui *Homo digitalis*, atât ceea ce este, cât şi locul în care acţionează vor fi în mare parte digitale. *Homo digitalis* va fi la început o versiune digitală a noastră. Pe măsură ce calculatoarele vor deveni mai inteligente, vom externaliza din ce în ce mai mult din gândirea noastră către ele. Aceste entităţi digitale nu vor mai fi ţinute în frâu de creierele noastre complexe, dezordonate şi oarecum limitate. Vom scăpa de limitele corpurilor care au nevoie de odihnă şi de somn şi care în cele din urmă se degradează şi mor. Nu vom mai fi limitaţi la a observa şi la a acţiona într-un

singur loc. Vom fi peste tot în mod simultan.

Prin augmentarea creierului nostru în mod digital, *Homo digitalis* va fi mult mai deştept decât *Homo sapiens.* Distincţia dintre ceea ce gândim noi şi ceea ce este gândit în cloudul IA va fi din ce în ce mai greu de făcut. *Homo digitalis* va transcende forma noastră fizică şi va fi o fiinţă atât biologică, cât şi digitală. Vom trăi în acelaşi timp şi în creierul nostru, şi în spaţiul digital, cu mult mai mare.

De fapt, în cea mai mare parte a timpului, *Homo digitalis* nu va trebui să facă parte din lumea analogă, care este înceată, dezordonată şi periculoasă. Vom trăi şi vom acţiona din ce în ce mai mult într-o lume pur digitală. După un secol de schimbări climatice, crize financiare şi terorism, această lume digitală va fi un loc primitor, organizat şi bine ordonat. Nu va mai exista nicio fărâmă din incertitudinea care face viaţa pe pământ atât de dureroasă uneori. Nu vor mai exista cutremure şi alunecări de teren. Nu vor mai exista epidemii. Totul va urma reguli precise şi corecte. *Homo digitalis* va fi stăpânul acestui univers digital. Vom deveni, într-un fel, zeii acestui spaţiu digital.

Acesta este rezultatul la modul optimist – pentru că vom ajunge să construim acest viitor digital. Chiar suntem nişte zei în sensul acesta. Şi ne putem asigura că acest viitor va fi corect, drept şi frumos. Sau putem

lăsa forţele care ne controlează planeta în prezent să definească felul în care va arăta acest sistem, şi astfel vom permite să fie plin de inegalitate, nedreptate şi suferinţă. Avem ocazia de a alege. Şi începem să luăm aceste decizii astăzi.

Viitorul nu este inevitabil. Este produsul deciziilor pe care le luăm azi. Dar mie mi se pare foarte probabil că ne aflăm într-un moment critic. Există mai multe forţe care ne împing spre o pantă alunecoasă, una care va conduce la o lume problematică şi groaznică.

Acum avem ocazia să facem nişte alegeri care ne vor salva de acest sfârşit şi care ne vor ghida spre un viitor digital mai luminos. Unele alegeri vor fi uşoare şi la îndemână, altele dificile şi costisitoare. S-ar putea să necesite lideri cu viziune, altruism, ba chiar sacrificiu.

Am fost foarte norocoşi până acum. Am fost stăpâni pe această planetă – acest punct minunat, albastru-verde, care se învârte în jurul unei stele destul de tipice de pe o spirală minoră din Calea Lactee – în ultimele sute de mii de ani. Le datorăm nepoţilor noştri – care la urma urmelor vor face parte din această nouă specie numită *Homo digitalis* – să facem în următoarele decenii ceea ce trebuie.

Cui se adresează această carte?

Această carte este scrisă pentru toţi cei care sunt îngrijoraţi cu privire la direcţia în care ne va aduce IA. Sunt multe aspecte la care trebuie să ne gândim. Va face IA să dispară slujbele oamenilor, inclusiv pe cele care necesită creativitate? Va deveni IA conştientă? Ce înseamnă IA în contextul conceptului de liber arbitru? Ce valori etice va avea (sau ar trebui să aibă) IA? IA va ajuta sau va afecta societatea? Va schimba felul în care ne vedem pe noi înşine? Va schimba esenţa umanităţii noastre?

Vei descoperi răspunsurile mele la aceste întrebări pe măsură ce voi discuta impactul social şi etic al trecerii noastre în lumea digitală. În mare parte, voi examina tendinţele pe care le putem observa azi, şi voi extrapola potrivit lor. Dar prezentul nu are un viitor fix. Alegerile pe care le facem acum şi în viitorul apropiat vor determina viitorul îndepărtat. Aşadar, voi prezenta atât variantele de viitor bune, cât şi unele mai puţin bune. Depinde de noi toţi să lucrăm astfel încât să ajungem la nişte rezultate bune.

Această carte se concentrează pe anul 2062. Aşa cum voi povesti în capitolul următor, majoritatea experţilor în IA cred că există o probabilitate de 50% că vom fi creat maşinării care pot gândi la fel de bine ca noi până în anul 2062. Este posibil ca această dată să fie puţin cam optimistă, şi va trebui să aşteptăm până în jurul

anului 2220 pentru ca IA să ajungă la nivel uman. Majoritatea experţilor consideră că există o certitudine de 90% că până atunci vom fi reuşit acest lucru. Oricare va fi această dată, magia începe abia atunci când maşinăriile vor depăşi nivelul nostru de inteligenţă.

Această carte este pentru cititorii interesaţi de, dar nu experţi în acest domeniu. Vor fi câteva grafice, dar nicio ecuaţie. Nu voi descrie ce este IA, şi nici ce a reuşit să obţină până acum – pentru aceste lucruri vă recomand să citiţi cartea mea anterioară *Its Alive.* În notele de subsol veţi găsi referinţe, explicaţii suplimentare şi câte o observaţie amuzantă, dar le puteţi ignora complet şi tot vă veţi putea bucura de această carte[12]. Dar dacă doriţi să explorаţi o idee tehnică mai în profunzime, aceste notiţe vă vor oferi mai multe detalii şi o trambulină către literatură.

Filosoful Nick Bostrom a prezis în anul 2015 că, „pe termen lung, inteligenţa artificială va fi o chestiune foarte importantă – probabil cel mai semnificativ lucru pe care l-a făcut umanitatea vreodată”.[13] Dacă are dreptate, noi trebuie să explorăm aceste consecinţe.

---

[12] Da, aşa cum v-am sfătuit şi în cartea mea anterioară, puteţi să ignoraţi complet notele de subsol!

[13] Robert Gebelhoff, „Q&A: Philosopher Nick Bostrom on Superintelligence, Human Enhancement and Existential Risk“, *The Washington Post*, 5 noiembrie 2015.

0002
Sfârşitul nostru

Am avut sute de ani la dispoziţie să ne obişnuim cu ideea că maşinăriile ar putea fi mai bune decât noi. În trecut, totuşi, numai forţa musculară ne era depăşită, deoarece maşinăriile puteau să depună mai mult efort fizic decât orice om. Dar în ultimii cincizeci de ani, creierele noastre sunt depăşite din ce în ce mai mult – cel puţin dacă ne concentrăm pe sarcini limitate din punct de vedere intelectual. Până în anul 2062 este posibil ca întrecerea să se încheie. *Homo digitalis* va câştiga.

Poate că te va surprinde faptul că primul campion mondial a fost învins de un calculator cu aproape patru decenii în urmă. Pe data de 15 iulie 1979, campionul mondial la table, Luigi Villa, a fost bătut în mod clar de programul BKG 9.8 al lui Hans Berliner, la un scor de 7-1. Ceea ce a făcut situaţia şi mai neplăcută pentru Villa a fost faptul că era campion mondial doar de o singură zi înainte să fie învins. Mai recent, în anul 1997, campionul la şah de la acea vreme, Garry Kasparov, a fost bătut la limită de calculatorul Deep Blue al corporaţiei IBM. Vorbind despre înfrângerea sa, Kasparov a descris un viitor pe care umanitatea îl avea în faţă:

Am mai jucat cu multe calculatoare, dar nu am mai simţit niciodată ceva de genul ăsta. Puteam să simt – puteam să miros – un tip nou de inteligenţă de partea cealaltă a mesei. În timp ce jucam restul jocului cât de bine am putut, m-am simţit pierdut; a jucat frumos, un joc de şah perfect, şi a câştigat cu uşurinţă.[14]

La fel ca înfrângerea lui Villa, pierderea lui Kasparov în faţa lui Deep Blue a fost un rezultat nemilos. Kasparov este considerat de mulţi oameni ca fiind unul dintre cei mai mari jucători de şah din toate timpurile. În anul 1985, când a devenit campionul mondial la şah pentru prima dată, era cel mai tânăr jucător care a ajuns să deţină acest titlu în cadrul acestui sport. Douăzeci de ani mai târziu, când s-a retras din şahul profesionist, Kasparov a rămas jucătorul cu cel mai mare calificativ. Pardon, a rămas jucătorul uman cu cel mai mare calificativ. Este trist pentru Kasparov că s-ar putea ca lumea să-şi aducă aminte de el ca fiind primul campion mondial la şah care a fost bătut de un calculator.

Programele de şah au progresat semnificativ din anul 1997. Nici Kasparov, nici actualul campion mondial, Magnus Carlsen, nu ar putea să facă faţă celor mai

---

[14] Garry Kasparov, „The Day That I Sensed a New Kind of Intelligence”, *Time*, 25 martie 1996.

bune programe din zilele noastre. Într-adevăr, Kasparov s-ar chinui să-l bată şi pe Pocket Fritz 4, care este un program pe telefon mobil. Pocket Fritz 4 are o evaluare Elo[15] de 2898, mai mare decât evaluarea maximă a lui Kasparov, de 2851.

Atunci când un program primeşte mai multe resurse informatice decât un telefon mobil, şansele de a-l învinge sunt prea mici. Deep Fritz, care operează pe un PC standard, are un scor Elo extraordinar, de 3150. Diferenţa de 300 de puncte dintre Kasparov şi Deep Fritz înseamnă că rusul are o şansă mai mică de una la cinci de a câştiga orice joc, şi aproape nicio şansă de a câştiga un turneu. Pentru cineva ca mine, cu o evaluare Elo mult mai mică, şansele mele de a câştiga un singur joc în faţa lui Deep Fritz sunt aproape nule.

Dar şahul uman nu a suferit din cauza dominaţiei de către computere. Într-adevăr, şahul pe calculator a îmbunătăţit şahul uman în mai multe feluri. Şahul pe

---

[15] Clasamentele Elo descriu nivelurile relative ale jucătorilor care joacă partide de doi jucători, cum ar fi în cazul şahului. Sistemul este numit după creatorul său, Arpad Elo, profesor de fizică ungaro-american. Clasamentul Elo al unui jucător este actualizat de fiecare dată când câştigă sau pierde, şi variază în funcţie de clasamentul Elo al adversarului său. Este puţin dificil să ai încredere totală în clasamentele Elo ale programelor de şah pe calculator, întrucât în general au jucat prea puţine jocuri în condiţii de campionat. în ciuda acestui lucru, diferenţa dintre cei mai buni oameni şi cele mai bune programe pe calculator este atât de mare acum, încât nu a mai rămas prea multă speranţă pentru oameni.

calculator acum oferă şi antrenamente profesioniste amatorilor umani. Şi calculatorul a deschis noi mijloace de joc pe care noi, oamenii, nu le-am fi luat niciodată în considerare. Stăpânii noştri roboţi au îmbunătăţit jocul uman.

Toate sistemele sunt Go

În martie 2016, a fost depăşit un alt punct de referinţă în istoria IA, când programul Alphago al Deepmind l-a bătut pe Lee Sedol, unul dintre cei mai buni jucători de Go de pe planetă. Go este un joc chinez antic şi foarte complex, care se joacă prin plasarea unor pietricele albe şi negre pe o tablă de 19 pe 19 pătrăţele, cu scopul de a câştiga un teritoriu cât mai mare.

Go este mult mai incitant decât şahul din mai multe motive. La şah, pot exista 20 de mişcări posibile de fiecare dată când unui jucător îi vine rândul. La Go, pot fi în jur de 200 de variante disponibile[16]. La şah, adesea nu este foarte greu să îţi dai seama cine câştigă: fiecare piesă de pe tablă poate primi un scor, şi cel mai probabil să bată este jucătorul cu cel mai mare scor. La

[16] La prima mutare într-un joc de Go, jucătorul alb poate să joace oricare dintre cele 361 de poziţii (19x19); la a doua mutare, jucătorul negru poate să joace oricare dintre cele 360 de poziţii rămase; la a treia, jucătorul alb poate să joace oricare dintre cele 359 de poziţii rămase; şi aşa mai departe.

Go, toate piesele sunt identice. Ca să decizi cine câştigă, este nevoie să te gândeşti mult mai subtil şi să evaluezi ce teritoriu este controlat de fiecare jucător. Unui om îi ia ani întregi să înveţe cum să joace Go bine.

În luna mai a anului 2017, Deepmind a demonstrat în mod convingător că rezultatul din 2016 al lui Alphago împotriva lui Sedol nu a fost o pură întâmplare. Într-un meci în valoare de 1,8 milioane de dolari[17], un Alphago îmbunătăţit l-a bătut pe copilul-minune chinez Ke Jie, care era pe atunci numărul unu mondial.

Dar deşi aceste două victorii au fost nişte momente-cheie pentru IA, semnificaţia lor a fost exagerată în anumite cercuri. Alphago a fost special antrenat ca să joace Go. Ar fi nevoie de mult efort pentru a adapta un program să joace un joc atât de diferit cum ar fi pokerul[18]. Nu prea este posibil ca

---

[17] Toate sumele de bani precizate în această carte, cu excepţia cazurilor în care specific altă valută, sunt în dolari americani.

[18] În luna octombrie a anului 2017, Deepmind a lansat Alphago Zero. Acesta a îmbunătăţit versiunea anterioară a Alphago, întrucât nu avea caracteristici antrenate de mână şi nici nu a fost expus la jocuri ale experţilor umani. Nu i s-au dat decât regulile jocului Go. Aşadar, nu a fost construit pe baza miilor de ani de cunoştinţe umane referitoare la Go, ci a învăţat totul de unul singur. După trei zile în care a jucat de unul singur, performa la un nivel suprauman. La fel ca mulţi dintre colegii mei în IA, am fost impresionat. Am fost chiar şi mai impresionat în decembrie 2017, când compania a lansat Alpha Zero, o versiune şi mai generală, care putea să înveţe să joace şah şi shogi (şah japonez) la nivel suprauman, numai cu ajutorul regulilor. Cu toate acestea, nu este clar (şi într-adevăr nu pare probabil, din punctul meu de vedere) că aceste programe ar

aceleaşi tehnici care funcţionează în cazul Alpha Zero (cea mai recentă versiune a Alphago, care este creată să funcţioneze pe baza regulilor oricărui joc cu tablă) să funcţioneze şi în cazul unui joc care implică noroc şi probabilităţi. Şi cu siguranţă Alpha Zero nu este deloc util pentru conducerea unei maşini autonome, pentru scrierea unui roman sau pentru traducerea unui document legal.

O altă concepţie greşită este că rezultatul Alphago a apărut dintr-odată şi, prin urmare, indică un fel de îmbunătăţire „exponenţială" în IA. Dar acest lucru nu este adevărat. A fost cu certitudine o realizare importantă care a captat atenţia şi imaginaţia oamenilor, iar Deepmind merită toate felicitările. Deşi Alphago a adus împreună nişte componente într-un fel nou, nu era de fapt nimic nou cu privire la aceste

---

putea învăţa să joace un alt tip de joc. Şahul, Go şi shogi sunt toate jocuri de tablă cu doi jucători. Pokerul, de exemplu, nu doar că introduce şi mai mulţi jucători, ci şi multe alte caracteristici, inclusiv nesiguranţa şi psihologia umană. Ca să câştigi la poker, trebuie să te descurci cu informaţiile incomplete despre cărţi, în timp ce la Go toate informaţiile despre stadiul jocului sunt valabile ambilor jucători. La poker, trebuie să te confrunţi şi cu trucurile psihologice ale adversarului tău, cum ar fi jocul la cacealma. Nici Alphago, nici Alpha Zero nu sunt create pentru a se descurca cu vreuna dintre aceste caracteristici. Ca să demonstreze independenţa în acest domeniu, Deepmind ar trebui să arate că acelaşi program ar putea să câştige o serie variată de jocuri, cum ar şi şahul, pokerul şi *Starcraft.* Chiar şi atunci, algoritmul Alpha Zero ar fi limitat doar la jocuri.

componente[19].

Cel mai de succes program pe calculator înainte de Alphago a fost Crazystone, creat de Remi Coulom. Într-un interviu din anul 2014, Coulom a prezis că va dura zece ani până să bată un jucător profesionist. Dar adevărul este că lui Alphago i-a luat puţin peste un an ca să-l învingă pe Fan Hui, care a fost campion european de trei ori, şi încă un an ca să-l învingă pe Lee Sedol.

Totuşi, Deepmind a investit mult mai mult efort în această chestiune, aşa cum nu s-a mai întâmplat niciodată. Programele Go erau create de persoane; Deepmind a avut peste cincizeci de oameni care lucrau la Alphago. A durat de zece ori mai puţin să se ajungă în acel punct, dar munca umană a fost de zece ori mai multă.

Deepmind a avut acces şi la fermele de servere vaste ale Google, lucru ce i-a permis lui Alphago să joace miliarde şi miliarde de partide de Go cu el însuşi. Dacă nu ai face nimic altceva decât să joci Go toată viaţa, tot

---

[19] AlphaGo nu a fost prima reţea neurală artificială care a învăţat cum să joace un joc la nivel uman. TD-Gammon a fost un program de table pe calculator dezvoltat în anul 1992 de Centrul de Cercetare Thomas J. Watson din cadrul IBM. TD-Gammon juca puţin sub nivelul celor mai buni jucători de table de la acea vreme. Explora strategii pe care oamenii nu le încercaseră, şi a condus la progrese în privinţa înţelegerii jocului de table. La fel ca AlphaGo Zero, a pornit de la regulile jocului şi a învăţat de unul singur cum să joace bine.

nu ai putea să ajungi la un număr atât de mare de partide. Aşa că Alphago este de fapt un elev care învaţă foarte încet. Spre deosebire de programe ca acesta, oamenii pot să înveţe din câteva exemple. Nu va fi uşor să construim un program IA care să poată învăţa din atât de puţine date. Aşa că, deşi victoriile lui Alphago au fost nişte momente simbolice pentru IA, probabil că nu a fost exact schimbarea pe care departamentul de relaţii cu publicul de la Google te-a făcut să crezi că a avut loc[20].

Dincolo de jocuri

Jocurile reprezintă o provocare simplă pentru IA. De obicei au reguli clare şi un câştigător evident. Cât despre jocurile precum şahul şi Go, în general se consideră că necesită un nivel rezonabil de inteligenţă din partea jucătorilor. Aşadar, nu este surprinzător că au fost testul perfect pentru cercetările în domeniul inteligenţei artificiale.

---

[20] Publicitatea pe care victoria lui AlphaGo a adus-o companiei Google pe valoroasa piaţă chineză a plătit probabil pentru milioanele de dolari pe care DeepMind i-a investit în AlphaGo. Dar acest lucru s-ar putea să se întoarcă împotriva celor de la Google, întrucât chinezii au fost motivaţi să dezvolte şi ei IA. Dacă unul dintre giganţii chinezi cum ar fi Baidu sau Tencent câştigă cursa IA în faţa lui Google, Larry Page şi Sergey Brin s-ar putea să regrete ziua în care a fost trezit la viaţă gigantul chinez.

Dar calculatoarele nu au devenit mai bune decât oamenii doar în domeniul jocurilor. Observăm cum într-un număr de domenii mai practice calculatoarele încep să îi surclaseze pe oameni. În domeniul medicinei, computerele sunt acum mai bune decât oamenii în anumite cazuri, cum ar fi atunci când vine vorba de electrocardiograme. O echipă de la Universitatea Stanford, condusă de Andrew Ng, fost şef al departamentului de cercetare în domeniul IA de la Baidu, a construit un model de IA care poate să identifice aritmia cardiacă dintr-o electrocardiogramă mai bine decât un expert uman.

Cancerul constituie un al doilea exemplu. O echipă de la Google a reuşit să folosească învăţarea automată pentru a detecta cancerul la sân din cadrul unor rapoarte anatomopatologice cu mai multă acurateţe decât doctorii umani. Şi face asta mult mai rapid – şi astfel mult mai ieftin – decât oamenii.

Un al treilea exemplu datează din anii ’80, când sistemul expert PUFF diagnostica bolile pulmonare într-un spital din California, alături de doctori oameni. Inteligenţa artificială ne oferă deja servicii medicale mai bune, mai rapide şi mai ieftine.

Şi în afaceri există multe domenii în care calculatoarele îi întrec pe oameni. Să luăm ca exemplu bursa de valori. Blackrock este cel mai mare gestionar

al banilor din lume. Se ocupă de peste cinci trilioane de dolari. Multe dintre fondurile sale administrate activ sunt acum întreţinute de algoritmi. Calculatorul are un avantaj în faţa managerilor de fonduri umani datorită volumului imens de date pe care le poate analiza. Calculatoarele pot să îndeplinească sarcini pe care niciun om nu le poate îndeplini, cum ar fi monitorizarea datelor prin satelit ce provin din parcările magazinelor sau din căutările de pe internet pentru a prevedea care vor fi volumul vânzărilor şi creşterea economică.

Asigurările sunt încă o arie în care calculatoarele îi depăşesc pe oameni. În Japonia, Fukuoku Mutual Life Insurance procesează acum plăţi folosind Watson, propunerea IA a IBM. Când a început să folosească Watson, IBM a pus pe liber treizeci şi patru de angajaţi care se ocupau de această sarcină. Compania se aşteaptă acum să economisească un milion de dolari în fiecare an.

Pentru al treilea exemplu, să revenim la legi. Un număr de afaceri noi, inclusiv Luminance, pot să proceseze automat seturi de date mari şi nestructurate pentru a-i ajuta pe avocaţi să se ocupe de verificarea prealabilă a contractelor. Software-ul poate să găsească anomalii de două ori mai repede decât înainte. De asemenea, reduce şi expertiza necesară îndeplinirii

unor astfel de sarcini.

Astfel de moduri de utilizare a IA deja transformă multe zone de lucru. Într-adevăr este greu să ne gândim la un sector al economiei care nu va fi afectat până în anul 2062.

Inteligenţa artificială generală

Toate sistemele de IA despre care am discutat până acum rezolvă o problemă limitată. Jocul Go. Citirea mamografiilor. Alegerea unor acţiuni. Scopul inteligenţei artificiale *generale* (IAG) este acela de a construi programe care pot face *orice* la fel de bine – sau chiar mai bine – ca oamenii. Încă suntem oarecum departe de IAG. Şi spre deosebire de toată vâlva despre care citeşti în presă în ziua de azi, există nişte obstacole substanţiale între stadiul în care ne aflăm acum şi IAG.

În primul rând, oamenii învaţă rapid. Trebuie să învăţăm rapid, este în ADN-ul nostru. Când eşti fugărit de un tigru, nu ai timp la dispoziţie ca să înveţi din greşeli. Sistemele IA, pe de altă parte, încă învaţă destul de încet. Reuşitele recente de învăţare profundă, în cadrul unor domenii cum ar fi jocul Go, transcrierea mandarinei şi recunoaşterea imaginilor au fost posibile datorită unei cantităţi mari de date.

Există multe situaţii în care nu avem multe date –

într-adevăr, sunt multe pentru care nu vom avea *niciodată* suficiente date. Spre exemplu, robotul meu se va distruge dacă va cădea de prea multe ori în timp ce învaţă să meargă. În mod similar, nu putem avea prea multe date despre problemele medicale rare. Sau despre cum se poate prevedea căderea burselor de valori. Ca să ne ocupăm de aceste situaţii, trebuie să construim sisteme IA care pot învăţa rapid, la fel ca oamenii.

În al doilea rând, oamenii se pricep foarte bine să-şi explice deciziile. Aceasta este încă o chestiune importantă când vine vorba de luarea deciziilor. Poate că nu voi accepta decizia doctorului să mă opereze dacă nu-mi poate explica de ce este necesară operaţia. Reactorul nuclear trebuie să explice de ce se închide. Dar sistemele IA încă sunt doar nişte cutii negre[21]. Oferă răspunsuri, dar nu pot să explice cum au ajuns la aceste răspunsuri. Un algoritm de învăţare profundă îţi poate spune că aceasta este poza unei pisici. Dar nu-ţi poate spune de ce este o pisică. Că este o pisică pentru că are blană, patru picioare şi lăbuţe drăgălaşe. Dar nu

[21] In industria aviaţiei, o cutie neagră (care de fapt nu este niciodată neagră - de obicei este roşie sau portocalie) înregistrează multe statistici interne legate de avion. în IA, o cutie neagră este un sistem în care poţi vedea doar datele de intrare şi cele de ieşire. Nu ai nicio informaţie despre stările interne, şi cum transformă aceste date de intrare în date de ieşire. Opusul unei cutii negre este o cutie de sticlă, care îţi oferă o vizualizare în procesele interne.

îţi poate spune de ce nu este un câine. Încă avem nevoie să construim sisteme IA care să îşi poată explica deciziile.

În al treilea rând, noi, oamenii, avem o înţelegere profundă a lumii în care trăim. Atunci când ne naştem, nu ştim aproape nimic despre lume şi despre felul în care funcţionează. Nu ştim că un măr cade pe jos din cauza gravitaţiei. Că ploaia este apă evaporată ce cade din cer. Că Pământul se învârte în jurul Soarelui, iar Luna în jurul Pământului. Şi nici că Luna este supusă aceleiaşi forţe gravitaţionale care atrage mărul spre pământ. Învăţăm toate aceste lucruri şi multe altele. Adunăm toate aceste informaţii şi le sintetizăm până obţinem nişte cunoştinţe profunde legate de felul în care funcţionează universul nostru.

Dar sistemele IA din prezent nu au această capacitate. Dacă îi ceri unui program să-ţi traducă „Bărbatul era însărcinat”, nu înţelege de ce această propoziţie este ciudată. Dacă îi arăţi o imagine cu cineva care dă drumul unui măr, nu are habar că acel măr va cădea pe pământ, cu o acceleraţie de 9,8 m/s2. Încă nu am dezvoltat sisteme IA care înţeleg lumea aşa cum o înţelegem noi, pe de-a întregul. Sisteme care au, spre exemplu, raţiune, cum avem noi.

În al patrulea rând, oamenii sunt foarte adaptabili. Dacă suntem aruncaţi într-o situaţie nouă, vom începe

să ne adaptăm şi să ne descurcăm. Când tancul cu oxigen de pe *Apollo 13* a explodat, lumea întreagă a stat cu sufletul la gură trei zile în timp ce astronauţii şi coordonatorii de zbor s-au adaptat la ceea ce părea imposibil şi astronauţii s-au întors în siguranţă pe Pământ. Această capacitate de adaptare ne-a ajutat să devenim specia dominantă atât pe planetă, cât şi, în cazul acesta, în afara ei.

Sistemele IA, pe de altă parte, sunt extrem de instabile. Dacă modifici problema un pic, tind să se blocheze. Şi nu într-un mod graţios. De fapt, există un subdomeniu în IA dedicat găsirii unor modalităţi de blocare a sistemelor IA. Care este cea mai mică modificare din cadrul unei imagini care împiedică algoritmul să recunoască semnul de stop? Care este cea mai diferită imagine pe care algoritmul o „recunoaşte" incorect? Încă nu am creat sisteme IA care să greşească elegant, aşa cum se întâmplă în cazul oamenilor.

Cât timp ne-a mai rămas?

Într-o serie de domenii limitate, calculatoarele i-au surclasat pe oameni. Dar mai avem o distanţă lungă de parcurs până să putem construi IAG. Când vom ajunge în acel punct? Şi cât timp le va lua atunci calculatoarelor să devină mai inteligente decât noi?

Aceasta va fi problema noastră? Sau a copiilor şi a nepoţilor noştri? Luând în considerare milioanele de ani care au fost necesare pentru ca inteligenţa umană să evolueze, poate că acel punct este mult mai îndepărtat? Probabil că vor mai trece câteva secole, sau chiar milenii? Sau poate că nici nu se va întâmpla?

La o conferinţă din anul 2017 din Asilomar despre viitorul IA, Andrew McAfee a făcut următoarea observaţie: „Oricine face predicţii despre orice are de-a face cu viitorul Inteligenţei Artificiale ori vă păcăleşte, ori se păcăleşte pe sine". Eu unul voi ignora această afirmaţie înţeleaptă şi voi încerca să fac nişte predicţii pline de încredere. De fapt, nu eu voi face predicţiile, ci voi pune mai mulţi experţi IA să le facă. Să sperăm că vom da de înţelepciune în această mulţime de oameni.

În luna ianuarie a anului 2017, le-am cerut unora dintre colegii mei, mai precis trei sute dintre ei, cu toţii fiind cercetători care lucrează în domeniul IA, să estimeze cât mai exact cât timp va mai dura până vom trece peste toate obstacolele ce stau în calea IAG. Şi ca să pun răspunsurile lor în perspectivă, am cerut şi opinia unor non-experţi, aproape cinci sute.

Non-experţii au fost cititorii unui articol de ştiri pe care l-am scris despre programul de poker IA, Libratus, care tocmai ce bătuse câţiva jucători de top. În încheierea articolului meu, le-am cerut cititorilor să

completeze un mic chestionar despre om contra robot. M-am aşteptat să existe nişte diferenţe între predicţiile experţilor şi cele ale non-experţilor. Am avut dreptate.

Având în vedere faptul că există destulă incertitudine în ceea ce priveşte timpul necesar pentru construirea unei inteligenţe artificiale de nivel uman, sondajul le-a cerut experţilor şi non-experţilor să ofere trei predicţii. Când va exista o probabilitate de 10% ca un calculator să poată îndeplini cea mai mare parte a profesiilor umane cel puţin la fel de bine ca un om obişnuit? Şi când va exista o probabilitate de 50%? Dar una de 90%? Aceste întrebări au fost puse şi în cadrul unui studiu realizat în anul 2012, ce a fost prezentat în cartea *Superintelligence* a lui Nick Bostrom.

Deoarece progresele în lumea IA au atras multă atenţie în ultimii ani, m-a interesat să văd dacă datele pe care le-au oferit oamenii au fost apropiate de cele din 2012. Sondajul lui Bostrom a fost unul dintre principalele argumente prin care susţinea că IA reprezintă o ameninţare existenţială pentru umanitate relativ iminentă. Dacă ne aşteptăm ca sosirea IAG să se producă chiar mai curând, poate că ar trebui să-i luăm şi mai în serios avertismentele.

Dar nu a fost aşa. Experţii care au participat la sondajul meu au fost mult mai rezervaţi în comparaţie cu non-experţii în privinţa provocărilor pe care le ridică

construirea unei inteligenţe de nivel uman. Pentru o probabilitate de 90% ca un calculator să fie la fel de bun ca un om, în medie, predicţia experţilor a fost anul 2112, în comparaţie cu doar 2060, predicţia non-experţilor[22].

Hollywoodul şi agitaţia din jurul IA din prezent pot explica diferenţa de o jumătate de secol. Adesea fac glume pe seama faptului că cel mai bun lucru pe care l-ar putea face IA, pentru a creşte în ochii publicului şi pentru a aplana temerile oamenilor, ar fi să îşi găsească un birou în Los Angeles.

Pentru probabilitatea de 50%, predicţia mediană a experţilor a fost anul 2062. De acolo vine şi titlul acestei cărţi: anul în care, în medie, colegii mei din domeniul IA se aşteaptă ca omenirea să fi construit maşinării care să fie la fel de capabile precum oamenii. Acestei estimări îi corespunde predicţia non-experţilor care a reieşit a fi anul 2039, cu două decenii mai devreme. Non-experţii au fost puţin mai optimişti decât Ray Kurzweil, futurolog şi directorul departamentului de inginerie de la Google, care a prezis că oamenii vor fi depăşiţi de calculatoare în jurul anului 2045.

---

[22] Predicţia mediană este anul până în care 50% din grupul întrebat a prezis că computerele vor fi la fel de bune ca oamenii. Predicţia medie a fost o perioadă infinită, în viitor: un procent mic din experţi şi din non-experţi a prezis că computerele nu vor fi niciodată pe măsura oamenilor.

În cele din urmă, pentru probabilitatea de 10%, predicţia mediană a experţilor a fost anul 2034. Non-experţii pe de altă parte au propus anul 2026. Adică în mai puţin de un deceniu. Este aproximativ de două ori mai repede decât au prezis experţii.

De ce sunt experţii mai puţin optimişti decât non-experţii? Una dintre problemele de percepţie în ceea ce priveşte IA este că oamenii văd sisteme complexe jucând jocuri precum şah şi Go şi, gândindu-se că aceste jocuri necesită multă inteligenţă, atribuie acestor sisteme toate celelalte abilităţi intelectuale pe care le avem noi, oamenii. În cazul şahului uman şi al jucătorilor de Go, aceasta este o ipoteză rezonabilă. Cel mai probabil un bun jucător de Go este o persoană inteligentă. Dar asta nu este valabil şi în cazul calculatoarelor. Şi este o distanţă ca de la cer până la pământ între a juca Go şi a face multe dintre celelalte lucruri pe care le pot face oamenii şi care necesită inteligenţă.

Opinia mea, care este şi opinia majorităţii colegilor mei, este că va mai trece cel puţin jumătate de secol înainte să putem vedea cum calculatoarele vor fi la fel de bune ca oamenii. Având în vedere faptul că va fi nevoie de progrese majore, şi este greu de prezis când vor avea loc aceste progrese, s-ar putea să dureze chiar şi un secol sau mai mult. Dacă aşa stau lucrurile, vei

putea dormi liniştit la noapte.

Singularitatea tehnologică

Unul dintre motivele pentru care se crede că în viitor calculatoarele vor ajunge să dobândească o inteligenţă de nivel uman sau chiar suprauman rapid este ideea periculos de seducătoare a singularităţii tehnologice. Această idee a pornit acum cincizeci de ani de la John von Neumann, unul dintre fondatorii informaticii, şi de la matematicianul şi criptograful de la Bletchley Park, I.J. Good. Mai recent, a fost o idee popularizată de autorul de science-fiction Vernor Vinge şi de futurologul Ray Kurzweil.

Singularitatea este un moment anticipat din istoria omenirii în care vom fi dezvoltat o maşinărie atât de inteligentă, încât să se poată reproiecta singură în mod repetat astfel încât să fie mai inteligentă. Ideea este că acesta ar fi un moment-cheie, iar inteligenţa mecanică ar începe dintr-odată să se îmbunătăţească în mod exponenţial, depăşind cu rapiditate inteligenţa umană cu mai multe ordine de mărime.

Odată ce vom ajunge la singularitatea tehnologică, nu vom mai fi cea mai inteligentă specie de pe pământ. Va fi cu certitudine un moment important în istoria noastră. Ne temem că s-ar putea să se întâmple atât de

repede, încât nu vom avea timp să monitorizăm şi să controlăm dezvoltarea superinteligenţei, şi că această superinteligenţă ar putea duce – intenţionat sau neintenţionat – la sfârşitul speciei umane.

Susţinătorii singularităţii tehnologice – care de fapt sunt de obicei futurologii şi filosofii, nu cercetătorii IA – se comportă ca şi cum singularitatea ar fi inevitabilă. Pentru ei, este o certitudine logică; singura întrebare este când se va petrece acest lucru. Dar eu, la fel ca mulţi alţi cercetători din domeniul IA, am nişte dubii mari cu privire la caracterul ei inevitabil.

De-a lungul a 50 de ani am învăţat cât de dificil este să construieşti sisteme de calculator chiar şi cu o inteligenţă modestă. Şi nu am construit niciodată un singur sistem computerizat care să poată să se autoîmbunătăţească în mod repetat. Într-adevăr, până şi cel mai inteligent sistem pe care îl ştim – creierul uman – a avut parte doar de nişte îmbunătăţiri modeste în ceea ce priveşte abilităţile sale cognitive. Spre exemplu, pentru majoritatea dintre noi este încă la fel de greu ca întotdeauna să învăţăm o a doua limbă. Iar mai nimic din ce am reuşit să înţelegem despre creier nu a făcut această sarcină mai uşoară. Din anul 1930 a avut loc o creştere semnificativă şi progresivă în privinţa scorului la testele de inteligenţă în mai multe părţi ale lumii. Acesta se numeşte efectul lui Flynn,

după cercetătorul James Flynn din Noua Zeelandă, care a muncit mult ca să identifice acest fenomen. Dar explicaţiile pentru acest concept s-au concentrat mai mult pe îmbunătăţirea nutriţiei, a îngrijirilor medicale şi a accesului la şcoală decât pe felul în care ne educăm copiii[23].

Există mai multe motive pentru care singularitatea tehnologică s-ar putea să nu se petreacă niciodată. Am discutat despre mai multe dintre ele în cartea mea anterioară. Cu toate acestea, se pare că ideea cum că singularitatea este inevitabilă nu îşi va pierde din popularitate. Având în vedere importanţa subiectului – s-ar putea să decidă destinul speciei umane mă voi apleca din nou asupra acestor argumente, în detaliu, şi privindu-le din perspectiva evoluţiilor recente din aceste dezbateri. De asemenea, voi introduce nişte argumente noi împotriva caracterului inevitabil al singularităţii tehnologice.

Câinii care gândesc mai rapid

Prima mea obiecţie împotriva presupusei inevitabilităţi a singularităţii este o idee ce a fost numită argumentul *câinelui care gândeşte rapid.* Deşi

[23] Există indicii obţinute din testele de IQ din armatele daneză şi americană care arată că, de fapt, scorul mediu al IQ-ului a început să scadă în ultima vreme.

viteza calculatoarelor pare să se fi stabilizat, tot procesează date din *ce* în ce mai repede. Reuşesc să facă acest lucru deoarece exploatează din ce în ce mai mult paralelism şi îndeplinesc mai multe sarcini în acelaşi timp, oarecum asemenea creierului nostru.

Presupunerea este că dacă vor reuşi să se gândească mai mult şi mai intens la probleme, computerele vor reuşi până la urmă să devină mai inteligente decât noi. Şi am beneficiat fără îndoială de creşterea exponenţială a puterii computerelor; telefonul inteligent din buzunarul tău este o dovadă. Dar probabil că doar viteza de procesare de una singură nu ne va aduce la singularitate.

Să presupunem că ai putea creşte viteza cu care gândeşte creierul câinelui tău. Un astfel de câine tot nu ar putea să vorbească cu tine, să joace şah sau să compună sonete. În primul rând, nu posedă un limbaj complex. Cel mai probabil, un câine care gândeşte rapid va rămâne tot un câine obişnuit. Va visa că fuge după veveriţe şi beţe. Poate că va avea aceste gânduri mai rapid, dar cel mai probabil nu vor fi mai complexe de atât. În mod similar, calculatoarele cu o viteză mai mare nu vor avea şi o inteligenţă mai mare.

Inteligenţa este produsul a multor factori. Ne ia ani întregi de experienţă ca să ne antrenăm intuiţia. Şi de-a lungul acestor ani de învăţare ne cizelăm şi abilitatea

de a extrapola: preluăm idei din situaţii vechi şi le aplicăm unor situaţii noi şi nemaiîntâlnite. Ne îmbunătăţim cunoştinţele practice, ceea ce ne ajută să ne adaptăm la circumstanţe noi. Astfel, inteligenţa noastră presupune mai mult decât gândirea rapidă pentru a rezolva o problemă.

Punctul critic

Al doilea argument împotriva caracterului inevitabil al singularităţii tehnologice este *antropocentrismul* acestei teorii. Susţinătorii singularităţii acordă o importanţă specială inteligenţei umane. Ei consideră că depăşirea inteligenţei umane va fi un punct critic. Calculatoarele vor reuşi atunci să se reproiecteze şi să se îmbunătăţească singure. Dar de ce este inteligenţa umană un punct atât de important ce trebuie depăşit?

Inteligenţa umană nu poate fi măsurată pe o scară unică şi lineară. Şi chiar dacă asta ar fi posibil, nu ar fi reprezentată de un singur punct, ci ar fi un spectru întreg de inteligenţe diferite. Într-o cameră plină cu oameni, unii sunt mai deştepţi decât alţii. Deci ce standard de inteligenţă umană ar trebui să depăşească un calculator? Pe cea mai deşteaptă persoană din cameră? Pe cea mai deşteaptă persoană din lume din prezent? Pe cea mai deşteaptă persoană care a trăit

vreodată? Pe cea mai deşteaptă care s-ar putea să trăiască în viitor? Ideea de a depăşi „inteligenţa umană" începe deja să scârţâie puţin.

Dar să lăsăm pentru moment deoparte aceste obiecţii. De ce este inteligenţa umană, orice ar însemna asta, *punctul critic* ce trebuie depăşit, după care inteligenţa mecanică va merge ca pe roate? Se pare că ipoteza ar fi că dacă suntem suficient de isteţi să construim o maşinărie care este mai deşteaptă decât noi, atunci această maşinărie trebuie să fie şi ea suficient de deşteaptă încât să poată construi la rândul său o maşinărie şi mai inteligentă. Şi aşa mai departe. Dar nu avem niciun motiv logic să credem că aşa stau lucrurile. Poate că o să reuşim să construim o maşinărie mai inteligentă decât noi. Dar poate că ea nu va reuşi neapărat să se îmbunătăţească singură.

Ar putea exista un anumit nivel de inteligenţă care va reprezenta un punct critic. Dar ar putea fi vorba despre orice nivel de inteligenţă. Nu pare posibil ca acel nivel critic să fie *mai jos* decât inteligenţă umană. Dacă ar fi aşa, probabil că am putea simula o astfel de IA chiar azi, şi am folosi această IA ea să construim una mai inteligentă şi, aşadar, am începe deja procesul de autoîmbunătăţire repetată.

Deci se pare că un eventual punct critic trebuie să se afle la nivelul sau mai sus de nivelul inteligenţei

umane. Într-adevăr, ar putea depăşi cu mult inteligenţa umană. Dar dacă trebuie să construim IA cu o inteligenţă mult mai mare decât a noastră, este posibil ca noi să nu fim suficient de deştepţi ca să creăm astfel de maşinării.

Dincolo de inteligenţă

Al treilea argument pe care îl am împotriva inevitabilităţii singularităţii tehnologice are de-a face cu *meta-inteligenţa.* Inteligenţa, aşa cum am spus şi mai sus, presupune mai multe abilităţi diferite. Include abilitatea de a percepe lumea, dar şi pe aceea de a gândi raţional cu privire la acea lume percepută. Dar mai include şi multe alte abilităţi, cum ar fi creativitatea.

Argumentul pentru inevitabilitatea singularităţii confundă două abilităţi diferite. Combină abilitatea de a duce la îndeplinire o sarcină cu abilitatea de a-ţi îmbunătăţi singur capacitatea de a îndeplini o sarcină. Putem construi maşinării inteligente care-şi perfecţionează abilitatea de a face anumite lucruri, şi care ajung să facă aceste lucruri mai bine decât oamenii. Baidu, de exemplu, a construit Deep Speech 2, un algoritm de învăţare automată care a învăţat cum să transcrie mandarina mai bine decât oamenii. Dar Deep Speech 2 nu şi-a îmbunătăţit capacitatea de a învăţa

sarcini. Acum îi ia la fel mult să transcrie mandarina ca întotdeauna. Abilitatea sa supraumană de a transcrie mandarina nu a dus la progrese ale algoritmului de bază de învăţare profundă. Spre deosebire de oameni, care devin mai buni la învăţat pe măsură ce îşi însuşesc lucruri noi, Deep Speech 2 nu învaţă mai repede pe măsură ce acumulează mai mult.

Îmbunătăţirile în cadrul algoritmilor de învăţare profundă s-au petrecut prin vechea metodă: oamenii s-au gândit mult şi bine la o problemă. Încă nu am construit maşinării care să se poată perfecţiona singure. Şi nu suntem siguri că vom face asta vreodată.

Randamente în scădere

Al patrulea argument al meu împotriva inevitabilităţii singularităţii tehnologice este blestemul *randamentelor descrescătoare.* Chiar dacă computerele sunt capabile să se perfecţioneze constant, s-ar putea să nu fim recompensaţi cu câştiguri mari. Am fost martori la diminuarea randamentelor în mai multe domenii ale eforturilor umane. Spre exemplu, am îmbunătăţit în mod repetat eficienţa carburantului la maşini, dar îmbunătăţirile sunt acum din ce în ce mai mici pe măsură ce maşinile devin din ce în ce mai eficiente.

Să presupunem, ca exemplu, că pornim de la o

maşinărie care are inteligenţa medie a unui om. Prin definiţie, asta ar însemna că are un IQ de 100. Şi să presupunem şi că, din moment ce a creşte acest nivel de inteligenţă devine din ce în ce mai dificil, fiecare generaţie de maşinării inteligente va avea un IQ care creşte cu 50% faţă de cât a crescut în generaţia anterioară. IQ-ul este un mijloc de măsurare a inteligenţei imperfect, dar să presupunem că ar fi unul satisfăcător. A doua generaţie de astfel de IA ar avea un IQ de 150 – destul de impresionant. Dar probabil nici măcar la fel de inteligent ca tine. A treia generaţie ar avea un IQ de 175, a patra de 187,5 şi tot aşa. Indiferent câte generaţii vor mai urma, IQ-ul acestor roboţi cu autoîmbunătăţire repetată nu va trece niciodată peste 200.

Chiar dacă mărim miza, tot s-ar putea să dăm peste aceleaşi limite. Să presupunem că fiecare generaţie ar creşte nu cu 50%, ci cu o cifră impresionantă de 90% faţă de cea anterioară. A doua generaţie ar avea un IQ de 190. A treia unul de 271. În acest moment vom ajunge la limita inteligenţei umane atestate. A patra generaţie ne va duce dincolo de inteligenţa umană, la un IQ de 343,9. Dar indiferent câte generaţii vor mai fi în viitor, IQ-ul acestor maşinării impresionante nu va trece niciodată de 1000. Vor fi cu siguranţă extrem, extrem de inteligente, dar nu va fi vorba de o creştere

explozivă.

Limitele inteligenţei

Al cincilea argument contra caracterului inevitabil al singularităţii tehnologice ţine de *limita inteligenţei.* Chiar dacă computerele reuşesc să se îmbunătăţească în mod repetat, s-ar putea să ne confruntăm cu nişte limite fundamentale. Multe alte domenii au aceste limite; de ce inteligenţa nu le-ar avea?

Ştiinţa este plină de limite. În fizică, nu poţi depăşi viteza luminii. În chimie, există limite în ceea ce priveşte viteza unei reacţii chimice. În biologie, par să existe nişte limite fundamentale care ne împiedică să prelungim viaţa umană mult peste 120 de ani. Sau să putem alerga un maraton în mai puţin de două ore. E posibil ca şi IA să se lovească de nişte limite fundamentale?

Dacă intri într-un cazino şi începi să joci la ruletă, nu contează cât de deştept eşti, nu vei bate casa. Ruleta este practic creată să fie împotriva ta. Nu poţi să fii mai inteligent decât persoana care decide să iasă din cazino. Calculatoarele pot să calculeze probabilităţi mult mai bine decât oamenii. Aşadar, ele pot să acţioneze mult mai raţional decât oamenii. Dar calcularea

probabilităţilor în numere cât mai exacte s-ar putea să nu te ajute să învingi natura. Cea mai bună decizie s-ar putea să fie una pe care până şi nişte calcule mai simple şi mai puţin precise ar putea să o ofere.

Complexitatea computaţională

Al şaselea argument împotriva inevitabilităţii singularităţii tehnologice ţine de *complexitatea computaţională*, teoria matematică ce descrie dificultatea cu care se rezolvă diferite probleme computaţionale. Dacă nu facem trecerea la computere care se bazează pe forme de calcul pe care încă nu le avem în prezent, e posibil ca nici măcar îmbunătăţirile exponenţiale să nu ne ajute, din cauza limitelor fundamentale a ce pot calculatoarele să facă.

Legea lui Moore – dublarea puterii de calcul care a fost observată o dată la doi ani – ne-a amăgit probabil pe mulţi dintre noi că progresul tehnologic va rezolva majoritatea provocărilor computaţionale[24]. Trăim în

[24] Legea lui Moore este numită după Gordon Moore, co-fondatorul Fairchild Semiconductor şi Intel. In anul 1965, a afirmat că până atunci avusese loc o dublare a numărului de tranzistori componenţi ai unui circuit integrat în fiecare an. în anul 1975, a rotunjit la fiecare doi ani. Legea lui Moore a fost validă timp de 50 de ani. Este mai puţin cunoscut faptul că a expirat oficial de câţiva ani. Ca oricare alt trend exponenţial din lumea reală, la un moment dat trebuie să se lovească de o barieră. In acest caz, de limite cuantice. International Technology Roadmap for Semiconductors este organizaţia profesională care, aşa cum

vremuri exponenţiale, iar progresul exponenţial în lumea calculatoarelor pare să promită faptul că trebuie doar să aşteptăm să apară suficiente generaţii de computere. În zece ani, calculatoarele vor fi de o mie de ori mai puternice decât sunt azi. În 20 de ani, vor fi de un milion de ori mai puternice. În 30, de un miliard de ori. Cu siguranţă, la un moment dat vom avea suficientă putere computaţională încât să facem orice dorim, nu-i aşa? Din păcate, acest lucru este departe de adevăr.

Informaticienii au dezvoltat o teorie valoroasă despre complexitatea computaţională. Aceasta descrie de câtă computaţie este nevoie pentru rezolvarea a diverse probleme într-un mod precis şi abstract. Teoria complexităţii computaţionale se aplică indiferent de tipul de calculator folosit. Nu contează dacă este un PC sau un Mac, un telefon sau un ceas inteligent. Un dispozitiv diferit ar putea schimba timpul de rulare – timpul care este necesar rezolvării unei probleme –, dar doar cu un factor constant. Dar pe noi ne interesează

---

sugerează şi numele, planifică traseul ca să se obţină legea lui Moore. In anul 2014, ITRS a declarat că obiectivele industriei nu vor mai fi o dublare la fiecare doi ani. Şi dacă asta nu face parte din planul companiilor majore de semiconductori, atunci putem fi siguri că nu se va întâmpla. Nu va cheltui nimeni miliarde de dolari necesari construirii următoarei generaţii de semiconductori esenţiali pentru micşorarea tranzistorilor. Interesant este faptul că acum scopurile Intel sunt de a reduce consumul de energie astfel încât să avem mai multă energie computaţională pe telefoanele mobile.

creşteri ale timpului de rulare care sunt mult mai mari decât nişte simpli factori constanţi. Poate că suntem în căutarea unor creşteri exponenţiale şi, aşa cum vom vedea, a ceva mai mare decât nişte creşteri exponenţiale.

Să presupunem că vrem să calculăm cel mai mare număr de pe o listă. Aceasta este o problemă de timp linear. Trebuie să treci prin întreaga listă. Acest lucru necesită un interval de timp proporţional cu mărimea imputului, a listei de numere. Dacă dublezi mărimea listei, va dura de două ori mai mult să se găsească cel mai mare număr. Dacă triplezi lista, va dura de trei ori mai mult.

Acum să ne gândim cum am putea pune lista în ordine, de la cel mai mic la cel mai mare. O metodă simplă ar fi să găsim cel mai mic lucru mai întâi; aşa cum am observat, pentru asta este nevoie de timp care este proporţional cu lungimea listei. Apoi găseşti al doilea cel mai mic lucru şi aşa mai departe. În total, durata timpului pe care trebuie să-l petrecem rezolvând această sarcină sau oricare alta creşte cu lungimea listei la pătrat. Dacă dublezi lungimea listei, va dura de patru ori mai mult. Dacă triplezi lista, va dura de nouă ori mai mult. Dacă cvadruplezi lungimea listei, va dura de şaisprezece ori mai mult. Asta nu sună bine. Dar calculele pot să se complice la un nivel

mult superior.

Există probleme computaţionale în care timpul de rulare creşte exponenţial odată cu mărimea aportului. Să luăm ca exemplu problema unui cuplu care divorţează şi care trebuie să-şi împartă bunurile în două subseturi, fiecare subset având aceeaşi valoare. O metodă simplă pentru a se găsi o diviziune în subseturi cu valoare egală ar fi să calculăm sumele tuturor subseturilor posibile de bunuri. Dacă vreun subset are o valoare egală cu jumătate din valoarea totală, s-a ajuns la diviziunea în două subseturi cu aceeaşi valoare. De fiecare dată când imputul – lista de lucruri – mai primeşte un bun, numărul de subseturi care trebuie luate în considerare se dublează, şi (în cel mai rău caz) timpul de rulare al algoritmului nostru se dublează.

Vestea bună este că o îmbunătăţire exponenţială în puterea de calcul ne va ajuta să rezolvăm astfel de probleme. Fiecare dublare din puterea de calcul ne va permite să rezolvăm problema cu un obiect în plus. Oricare ar fi mărimea imputului pe care vrem să-l procesăm, în cele din urmă va fi în puterea noastră. Dacă vrem să procesăm liste ce conţin de zece ori mai multe lucruri decât ce este posibil în ziua de azi, mai trebuie să aşteptăm încă zece generaţii de calculatoare.

Dar mai sunt şi probleme computaţionale în care

timpul de rulare creşte mai rapid de atât. În cazul acesta, un progres exponenţial al puterii de calcul ar deveni insuficient pentru rezolvarea unor probleme mai mari. Să luăm în considerare problema calculului ariei mulţimii lui Mandelbrot. Mulţimea lui Mandelbrot este o figură fractală superbă ce conţine spirale şi forme de cai de mare extraordinare. A fost numită cea mai complexă figură a matematicii.

Ştim că mulţimea lui Mandelbrot are o arie finită. Se află într-un cerc cu o rază de 2, aşa că aria sa este de mai puţin de 4π (=12,566...). Dar calculul ariei precise este, din câte ştim noi, o provocare foarte mare din punct de vedere computaţional. Cea mai bună metodă pe care o avem pentru calcularea ariei converge foarte încet. Ar trebui să aduni 10^{118} termeni ca să ajungi să calculezi aria cu o precizie de până la două cifre, şi 10^{1181} ca să obţii trei cifre. 10^{118} este mult mai mult decât numărul de atomi din univers. Îmbunătăţirile exponenţiale ar putea să fie insuficiente în cazul unei probleme de calcul atât de dificile.

Bucle de feedback

Al şaptelea argument pe care îl am contra inevitabilităţii singularităţii tehnologice este că s-ar putea să se producă nişte reacţii nedorite care vor

împiedica apariţia singularităţii. Aceste bucle de feedback ar putea fi economice sau legate de mediu.

Martin Ford oferă un astfel de argument bazându-se pe feedbackul economic[25]. Înainte ca singularitatea să se producă, computerele vor fi suficient de capabile încât majoritatea locurilor de muncă să fie automate. Dar acest lucru ar duce la o rată de şomaj masivă. Şi în lipsa unor reforme radicale ale capitalismului, un astfel de şomaj ar face cererea de bunuri de consum să intre în colaps. Economia va fi distrusă şi nu se va mai putea investi în cercetare şi în dezvoltarea necesară pentru apariţia singularităţii.

Feedbackul ar putea ţine şi de mediu. Jared Diamond a susţinut că societăţile tind să se autolimiteze, dacă nu chiar să intre în colaps[26]. Succesul lor conduce rapid la depăşirea limitelor mediului lor. Colapsul urmează la scurt timp după succes. În cazul IA, bogăţia în creştere şi prosperitatea pe care le aduce ar putea să depăşească abilitatea mediului nostru de a susţine societatea umană. Drept consecinţă, singularitatea tehnologică nu s-ar mai putea realiza, pur şi simplu pentru că societatea se va prăbuşi din cauza unui exces de

---

[25] Vezi Martin Ford, *The Lights in the Tunnel: Automation, Accelerating Technology and the Economy of the Future*, Acculant Publishing, SUA, 2009.

[26] Vezi Jared Diamond, *Collapse: How Societies Choose to Fall or Succeed*, Viking Press, New York, 2005.

consum.

Frâne pentru inteligenţă

Al optulea argument împotriva inevitabilităţii singularităţii tehnologice vine de la Paul Allen, co-fondatorul Microsoft. El i-a dat numele „frâna de complexitate”. Cu cât progresăm mai mult în înţelegerea inteligenţei, cu atât devine mai dificil să facem progrese. Avem nevoie de din ce în ce mai multe cunoştinţe specializate, şi suntem forţaţi să dezvoltăm din ce în ce mai multe teorii ştiinţifice complexe. O astfel de frână de complexitate încetineşte progresul şi previne explozia inteligenţei artificiale. Allen a scris astfel:

Complexitatea extraordinară a cunoştinţelor umane ar trebui să-i facă precauţi pe cei care susţin că singularitatea se va întâmpla în viitorul apropiat. Dacă nu avem o înţelegere profund ştiinţifică a cogniţiei, nu putem crea un software care să declanşeze singularitatea. În loc de progresul care accelerează mereu, prezis de Kurzweil, noi credem că progresul către această înţelegere este încetinit fundamental de

frâna de complexitate.[27]

Allen observă că nu vom putea ajunge la inteligenţa artificială doar prin dezvoltarea unui hardware de computer mai rapid. Va fi nevoie şi de progrese semnificative în software. Şi cel mai probabil, pentru astfel de îmbunătăţiri va fi nevoie de descoperiri majore în privinţa înţelegerii cogniţiei umane. Aici intervine frâna de complexitate. Cogniţia umană se dovedeşte a fi o problemă deosebit de dificilă.

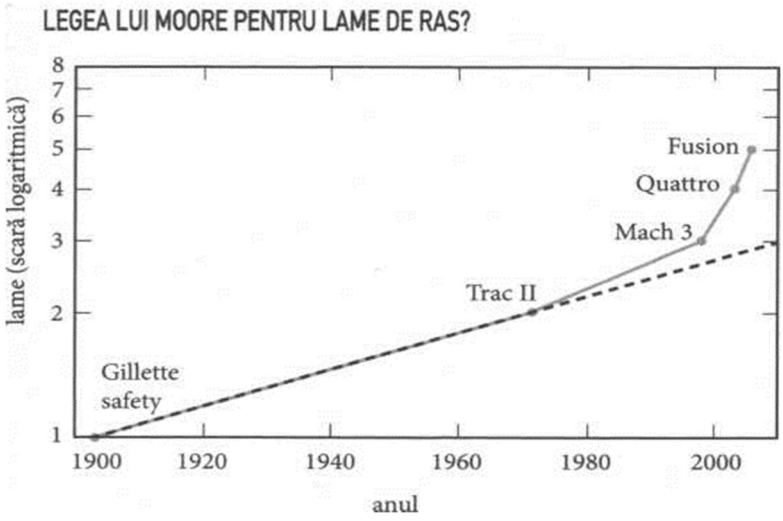
Extrapolare cu atenţie

Al nouălea argument împotriva inevitabilităţii singularităţii tehnologice este că trebuie să fim foarte atenţi la oamenii care extrapolează din grafice – în special grafice care au axe logaritmice. Revista *The Economist* a demonstrat acest lucru prin intermediul unui exemplu amuzant: umilul aparat de ras de unică folosinţă[28]. Poate că nu ai observat acest lucru, dar lamele de ras au crescut la număr exponenţial. Permite-mi să demonstrez prin acest grafic simplu.

---

[27] Paul Allen, „The Singularity Isn't Near“, *MIT Technology Review*, 12 octombrie 2011.

[28] „The Cutting Edge: A Moores Law for Razor Blades? “ The Economist, 16 martie 2006.



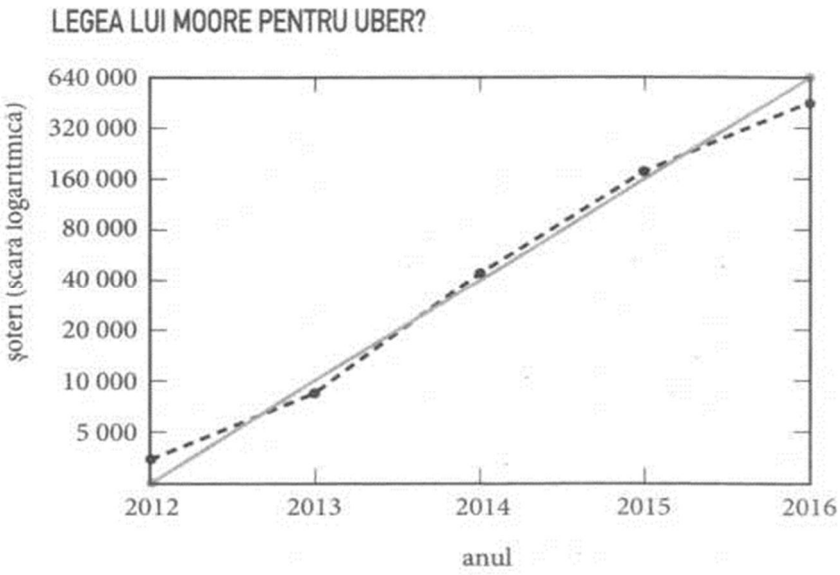
Metoda folosită de obicei şi testată prin care se demonstrează creşterea exponenţială este realizarea unui grafic cu o scală logaritmică verticală: 1, 2, 4, 8 şi aşa mai departe. La o astfel de scală comprimată, creşterea exponenţială devine o linie dreaptă simplă.

Linia punctată din grafic reprezintă creşterea exponenţială a numărului de lame de pe un aparat de ras de unică folosinţă. Asta obţii dacă numărul de lame se dublează la fiecare 68 de ani. În realitate, poţi observa că numărul de lame a crescut mult mai repede decât exponenţial. O lamă în anul 1903. Două lame 68 de ani mai târziu, în anul 1971. O altă dublare şi se ajunge la patru lame după doar 32 de ani, în 2003. Dacă extrapolăm din primii o sută de ani de tehnologie a aparatelor de ras de unică folosinţă, te-ai aştepta ca

până acum să fi rămas cu aparate cu două lame – nu cinci, aşa cum avem în cazul de faţă. Deci am avut o lege a lui Moore pentru aparate de ras. Dar să nu te aştepţi ca singularitatea bărbieritului să se petreacă prea curând.

Poate te vei plânge că te-am păcălit pentru că am folosit numere mici. Nu a fost greu să ajungem la o creştere exponenţială deoarece am trecut doar de la aparatul de ras Gillette cu o singură lamă din anul 1903 la aparatul Fusion cu cinci lame, din anul 2006. Aşa că permite-mi să-ţi ofer un exemplu în care numerele nu sunt mici.

Să ne gândim la numărul de şoferi de Uber de pe planetă. Uber îi numeşte „şoferi-parteneri”, dar adevărul este că sunt orice altceva numai parteneri în afaceri nu. Din nou, ca să demonstrăm creşterea exponenţială, voi crea un grafic cu o scară logaritmică ce restrânge axa verticală. Fiecare punct de pe axa verticală reprezintă un număr de şoferi Uber: 5000, 10.000, 20.000, 40.000, 80.000, 160.000 şi aşa mai departe.



La fel şi aici, linia punctată indică creşterea exponenţială. Vedem că s-a putut aplica legea lui Moore şi pentru numărul şoferilor de Uber. În fiecare an de când a început afacerea, numărul de şoferi de Uber aproape că s-a cvadruplat. Existau 8500 de şoferi de Uber în anul 2013. Numărul s-a cvadruplat şi a ajuns la 45.000 de şoferi în anul 2014. Şi o cvadruplare a avut loc din nou în anul 2015, când s-a ajuns la 180.000 de şoferi[29]. Dar asta nu înseamnă că vom avea vreun fel de singularitate Uber, nu se va ajunge în punctul în care toată lumea de pe planetă va conduce pentru Uber. În cele din urmă, această creştere exponenţială nu va putea fi susţinută. Numărul de şoferi Uber va trebui să

[29] [29] Jonathan Hall, „Taking Another Look at the Labor Market for Uber's Driver-Partners“, *Medium*, 22 noiembrie 2016.

stagneze la un moment dat. Vom ajunge în punctul de saturaţie a pieţei.

Un simplu model „viral” explică de ce, cel puţin iniţial, numărul de şoferi de Uber s-ar putea să se cvadrupleze anual. Să presupunem că, la fiecare şase luni, fiecare şofer încurajează un prieten să înceapă să conducă pentru firma Uber. Poate compania le oferă un bonus celor care fac asta. Exact de aşa ceva este nevoie pentru a fi stimulată creşterea exponenţială.

Să presupunem că Uber porneşte la începutul anului cu 10.000 de şoferi. Apoi aceşti 10.000 aduc şi ei încă 10.000 în următoarele şase luni. Se ajunge la 20.000. Aceşti 20.000 de şoferi aduc încă 20.000 de şoferi noi în următoarele şase luni. Rezultatul este că până la începutul celui de-al doilea an, Uber are 40.000 de şoferi. În mod similar, la începutul celui de-al treilea an, Uber va avea 160.000 de şoferi; şi 640.000 de şoferi la începutul celui de-al patrulea an. (Cam astea sunt numerele atinse în realitate.)

La fel ca orice schemă piramidală, o astfel de creştere exponenţială trebuie să se oprească până la urmă. Şoferii Uber rămân fără prieteni care nu sunt deja şoferi Uber. Alte companii de taxi rivale încep să ofere condiţii de muncă mai atractive decât Uber – şi asta nu e un lucru prea greu de făcut... Oricum, din mai multe motive, piramida se prăbuşeşte.

Se pare că sunt mai multe seturi de statistici pe care le putem măsura şi care par exponenţiale pentru o perioadă: numărul de lame dintr-un aparat de ras de unică folosinţă, numărul de şoferi de Uber, până şi costul mediu al unei educaţii mai bune. Dar asta nu înseamnă că se va ajunge la vreun fel de singularitate. Există multe legi, atât ale fizicii, cât şi ale economiei, care previn continuarea modelelor exponenţiale.

Să învăţăm din trecut

Ajung astfel la al zecelea şi ultimul argument împotriva inevitabilităţii singularităţii tehnologice. Trebuie să învăţăm din greşelile istoricilor din trecut. De fapt, trebuie să învăţăm din greşelile din trecut ale unuia dintre primii susţinători ai singularităţii.

Henry Adams a fost unul dintre primii oameni care au scris despre ideea singularităţii tehnologice. A fost nepotul lui John Quincy Adams, cel de-al şaselea preşedinte al Statelor Unite ale Americii, şi strănepotul lui John Adams, al doilea preşedinte şi unul dintre Părinţii Fondatori. În anul 1904, Henry Adams a publicat o autobiografie, *The Education of Henry Adams* („Educaţia mea”), care prezenta debutul secolului XX din perspectiva tinereţii sale, cu 50 de ani în urmă. Cartea a câştigat premiul Pulitzer şi este inclusă pe lista

celor mai importante cărţi de nonficţiune ale secolului XX.

Capitolul 34, „Legea accelerării”, propune o „lege a accelerării, precisă şi constantă ca oricare altă lege a mecanicii, [care] nu ne putem aştepta să îşi relaxeze energia pentru a se potrivi cu cerinţele omului”. Bazându-se pe această lege, Adams a susţinut că ştiinţa şi cunoştinţele se vor extinde atât de mult, încât societatea din anul 2000 va fi de neînchipuit pentru oamenii din anul 1900. Pentru a ilustra legea, Adams a prezentat faptul că din zece în zece ani, între 1840 şi 1900, cantitatea de energie eliberată per tonă de cărbune s-a dublat. Şi astfel, el a prezis că „fiecare american care va trăi în anul 2000 va şti cum să controleze o putere nelimitată”. Măcar de-ar fi aşa!

Desigur, Adams s-a înşelat. Nu am avut putere nelimitată în anul 2000. Iar anul 2000 nu a fost chiar de neimaginat pentru oamenii din 1900; mulţi au prezis destul de bine cum va arăta viaţa. H.G. Wells, spre exemplu, a prevăzut că se vor inventa tehnologii noi complexe, cum ar fi laserele şi armele nucleare[30]. Şi multe tendinţe pe care le vedem în ziua de azi, cum ar fi industrializarea şi globalizarea, erau deja destul de vizibile încă din anul 1900.

---

[30] H.G. Wells a prezis laserul în *Războiul lumilor* (1897) şi armele nucleare în *Lumea eliberată* (1914).

Maşinării superinteligente

Cele zece motive pe care le-am oferit pentru care este posibil să nu apară singularitatea nu demonstrează faptul că nu ar *putea* să apară. Acest lucru rămâne o posibilitate. Dar sper că te-am convins că nu este ceva inevitabil. Ca mulţi dintre colegii mei care au încercat să construiască computere care să se comporte cât de cât inteligent, mă îndoiesc de faptul că vom construi maşinării care să se îmbunătăţească singure constant, cel puţin prea curând. În acest caz, singularitatea tehnologică va rămâne în continuare o idee interesantă, dar una fictivă.

Deşi rămân sceptic cu privire la singularitatea tehnologică, asta nu înseamnă că nu vom ajunge la un nivel de inteligenţă umană sau chiar supraumană pentru computere. Chiar cred că vom ajunge în acest punct. De fapt, mi-am petrecut toată viaţa de adult lucrând pentru a atinge acest scop. Nu e nimic atât de special la inteligenţa noastră, şi nici la biologia noastră. Şi pentru acest motiv cred că într-o zi vom crea maşinării care gândesc. S-ar putea să gândească diferit faţă de noi[31]. Dar este posibil să ajungă la fel de

[31] Respectatul informatician savant Edsger Dijkstra (1930-2002) a spus cândva: „Întrebarea «Pot maşinile să gândească? »... este la fel de relevantă ca întrebarea

inteligente ca noi. Iar dacă ele ajung la inteligenţă de nivel uman, este dificil să ne imaginăm motive pentru care nu ne-ar depăşi, în cele din urmă. Deja au făcut asta într-un număr de domenii limitate. În plus, computerele au o serie de avantaje inerente în faţa noastră.

Dar nu-mi închipui că vom putea ajunge la superinteligenţă doar stând şi aşteptând să se îmbunătăţească ele singure. Presupun că vom ajunge la superinteligenţă la fel cum am obţinut toate celelalte progrese tehnologice din cursul vieţilor noastre: prin muncă grea, prin dezvoltarea ştiinţei graţie ingeniozităţii şi prin inginerie laborioasă pentru a face maşinăriile să fie mai deştepte decât noi.

Alphago a învăţat să joace Go mai bine decât orice om. Dar nu şi-a îmbunătăţit felul în care învăţa pe parcurs. Dacă i-ai cere lui Alphago să joace Go pe o tablă de începători, cu nouă pe nouă pătrăţele, în loc de obişnuitele 19 pe 19, ar trebui să înveţe din nou, aproape de la zero. Încă nu ştim cum să transferăm ce a învăţat pe o tablă mai mare la jocul cu o tablă mai mică. Un jucător uman de Go, pe de altă parte, ar putea începe să joace competent imediat pe o tablă de nouă pe nouă.

«Pot submarinele să înoate? »* (Conferinţa ACM South Central Regional, noiembrie 1984, Austin, Texas).

IA mai are mult până să ajungă să aibă inteligenţă de nivel uman, darămite superinteligenţă. Dar sunt de părere că vom ajunge acolo. Vom inventa computere care ne vor fi superioare. Vor fi mai puternice, mai rapide şi mai inteligente decât noi. Însă în loc să ne înlocuiască, sper că vom găsi o soluţie să facem ca aceste maşinării să ne amelioreze şi să ne extindă capacităţile.

Ameninţări existenţiale

Hollywoodul ne-a pregătit să fim îngrijoraţi că roboţi malefici o să încerce să preia controlul asupra planetei. Într-adevăr, potrivit filmului, Terminator va porni prima dată în anul 2029, aşa că ne-a mai rămas cam un deceniu. Oare Stephen Hawking a avut dreptate când a sugerat că superinteligenţă este cea mai mare ameninţare existenţială a noastră? În momentul de faţă, noi suntem cea mai inteligentă specie de pe pământ, şi celelalte vieţuitoare depind de bunătatea noastră ca să le permitem să continue să existe. Oare soarta noastră nu va depinde şi ea de bunătatea acestor computere superioare, superinteligente? Hai să ne concentrăm mai departe pe întrebarea: chiar ne vor elimina computerele superinteligente aşa, cu una, cu două? Vor reprezenta sfârşitul umanităţii?

În filme, IA sunt de obicei prezentate ca fiind rele. Dar mai degrabă incompetenţa, şi nu reaua-voinţă, pare cel mai probabil risc. Bineînţeles, trebuie să luăm în considerare posibilitatea că maşinăriile superinteligente ar putea să pună capăt speciei umane *neintenţionat.* Şi sunt mai multe scenarii în care s-ar putea întâmpla asta.

Atingerea lui Midas

În primul scenariu cu privire la riscurile puse de inteligenţa artificială, scopurile unui computer superinteligent ar putea fi proiectate prost. Acest lucru este ilustrat de mitul grecesc al regelui Midas, căruia i-a fost îndeplinită dorinţa de a transforma tot ce atingea în aur. Dar regele nu a specificat cum trebuie ceea ce-şi dorea: nu voia ca mâncarea şi fiica lui să fie transformate în aur.

Se poate spune că IA a trecut deja prin ceva de genul acesta, la o scară mai mică şi în moduri mai puţin periculoase. Spre exemplu, cercetătorii au raportat un experiment în anul 2017 în care au învăţat un calculator cum să joace *Coast Runners*, un joc video cu întreceri de bărci. În loc să termine cursa, IA a învăţat cum să se învârtă în cercuri mici, lovind alte bărci, deoarece acest lucru îi creştea scorul mai rapid decât dacă termina

cursa.

Fiind atât de deştepte, maşinăriile superinteligente ar putea să ne surprindă prin felul în care îşi ating scopurile. Să presupunem că îi cerem unei maşinării superinteligente să vindece cancerul. O cale prin care ar putea face asta ar fi să elimine toate gazdele care ar putea adăposti cancerul – punând astfel capăt speciei umane. Nu e exact ce ne-am dorit noi.

Astfel de exemple presupun o perspectivă negativă asupra superinteligenţei. Dacă eu ţi-aş da sarcina să vindeci cancerul şi tu ai începe să omori oameni, probabil că aş decide că nu eşti chiar atât de inteligent. Presupunem în general că oamenii isteţi au învăţat nişte valori morale şi sunt atenţi când vine vorba de suferinţa celorlalţi, mai ales a celor care au conştiinţă şi sentimente. Nu ar trebui o superinteligenţă să fie şi destul de înţeleaptă, nu doar inteligentă?

Agrafe cât vezi cu ochii

Un al doilea scenariu cu privire la riscurile IA este situaţia în care, în ciuda faptului că scopurile acesteia sunt specificate corect, ar exista nişte efecte adverse nedorite care ar răni umanitatea. Oricine a depanat coduri de calculator ştie cât de frustrant de literale pot să fie calculatoarele atunci când interpretează

instrucţiuni. Acest risc este explorat printr-un experiment mental foarte cunoscut, care a fost propus de Nick Bostrom.

Să presupunem că avem un robot superinteligent şi îi trasăm sarcina de a crea cât mai multe agrafe posibil. Întrucât robotul este superinteligent, se va descurca foarte bine să facă agrafe. Robotul ar putea începe să construiască din ce în ce mai multe fabrici de agrafe. În cele din urmă, toată planeta va fi transformată în fabrici de agrafe. Robotul face exact ce i s-a cerut să facă, dar rezultatul nu este unul dorit de oameni.

Bostrom nu crede de fapt că îi vom da unei maşinării scopul de a crea cât mai multe agrafe, în special pentru că suntem conştienţi de riscurile pe care le-ar reprezenta acest obiectiv. Exemplul cu producţia de agrafe a fost ales pentru a demonstra că până şi un scop obişnuit, arbitrar şi aparent inofensiv, poate să facă lucrurile s-o ia complet razna.

La fel ca argumentul cu Midas, acest exemplu include o perspectivă destul de pesimistă cu privire la superinteligenţă. Nu ar trebui o superinteligenţă să fie capabilă să înţeleagă scopurile implicite care nu sunt specificate explicit? Da, trebuie să producă multe agrafe, dar nu în detrimentul mediului înconjurător. Şi, cu siguranţă, nu în detrimentul speciei umane.

Un al treilea scenariu de risc presupune că superinteligenţa va avea scopuri secundare care ar putea intra în conflict cu existenţa continuă a umanităţii. Să presupunem că superinteligenţa are în mare un scop anume, cum ar fi fericirea omenirii sau protejarea planetei. Aproape oricare alt obiectiv de genul acesta pe care ţi-l poţi imagina va face necesar ca superinteligenţa să-şi dezvolte resursele necesare pentru a-şi duce la îndeplinire sarcinile. De asemenea, va trebui ca superinteligenţa să aibă permisiunea de a continua să opereze, astfel încât să-şi atingă scopul.

Dar oamenii ar putea să oprească maşinăria. Pe lângă asta, oamenii vor consuma resurse care ar putea fi folosite mai bine pentru atingerea obiectivelor superinteligenţei. Concluzia logică este că superinteligenţa ar vrea să ne elimine. Atunci nu am mai putea să o oprim sau să consumăm resurse care ar putea să fie folosite mai bine pentru scopurile ei.

Aceste scopuri secundare de autoconservare şi de achiziţie a resurselor au fost numite două dintre „impulsurile de bază ale IA" de Stephen Omohundro[32].

---

[32] Vezi Stephen Omohundro „The Basic IA Drives", în Pei Wang, Ben Goertzel & Stan Franklin (editori), *Artificial General Intelligence 2008: Proceedings of the First*

Aceste impulsuri sunt sub-scopurile de bază pe care orice IA suficient de inteligentă ar trebui să le aibă. Calculatorul HAL 9000 din filmul *2001: O odisee spaţială* al lui Arthur C. Clarke reprezintă probabil cea mai cunoscută viziune a impulsului IA de autoconservare. HAL începe să-i ucidă pe astronauţii de la bordul rachetei *Discovery One* într-o încercare disperată de a-i împiedica să-l stingă.

Alte impulsuri de bază ale IA ţin de îmbunătăţiri şi de creativitate. Sistemele IA vor tinde să devină mai eficiente, atât fizic, cât şi computaţional, deoarece aceste evoluţii le vor ajuta să atingă orice scopuri ar putea avea. Şi un lucru care este mai puţin previzibil, sistemele IA vor avea tendinţa de a fi creative, vor căuta mijloace noi prin care să-şi atingă ţelurile mai eficient şi mai direct.

Eficienţa nu este un lucru rău; ne va ajuta să conservăm resursele limitate ale planetei. Dar creativitatea presupune o provocare mai mare. Înseamnă că maşinăriile superinteligente vor fi imprevizibile. Îşi vor atinge obiectivele pe căi pe care noi nu le putem prezice. Acest risc va fi discutat mai detaliat în secţiunea următoare.

AGI Conference, Frontiers in Artificial Intelligence and Application 171, Amsterdam, IOS Press, 2008, pp. 483-492.

Un al patrulea scenariu ar fi că superinteligenţa s-ar putea modifica de una singură, ar putea începe să funcţioneze diferit sau chiar să-şi stabilească noi scopuri. Este posibil ca acest lucru să se întâmple dacă îi oferim obiectivul de a se face singură mai inteligentă. Cum putem fi siguri că superinteligenţa reproiectată de către ea însăşi va ţine cont de valorile noastre umane? Un aspect inofensiv al sistemului original ar putea fi amplificat în cel nou şi ar putea să devină periculos pentru noi.

Ţinta în mişcare ar putea fi nu doar superinteligenţa, ci şi sistemul mai mare în cadrul căruia operează. Observăm acest fenomen în instituţiile noastre umane: poartă numele de „misiune pe furiş". Iei decizia de a trimite nişte consilieri militari în Vietnam, şi un deceniu mai târziu te trezeşti că ai sute de mii de soldaţi pe terenul de luptă, purtând un război la scară largă ce nu poate fi câştigat.

Într-un fel modest şi destul de inofensiv, această problemă a ţintei în mişcare a fost deja observată în contextul IA. În anul 2008, Google a lansat Google Flu Trends. A devenit un simbol pentru folosirea *big data*[33]

[33] Termenul se referă la extragerea, manipularea şi analiza unor seturi de date care sunt prea mari pentru a fi tratate în mod obişnuit. Din această cauză se

pentru binele comun, prezicând mai eficient decât metodele anterioare perioada în care va fi sezonul de gripă în lume. Google Flu Trends a folosit căutările de pe Google Search pentru a prezice când şi unde va fi mai prevalentă gripa. Dacă mulţi oameni dintr-o anumită regiune încep să întrebe pe Google „Cum se tratează durerile în gât?" sau „Ce este febra?", atunci probabil că gripa începe să se răspândească. Dar în anul 2013, Google Flu Trends pur şi simplu nu a mai funcţionat. A fost scos pe tăcute de pe lista serviciilor oferite de Google. Ce a mers prost? Problema a fost că Google Search (şi ecosistemul uman în care se regăseşte) este o ţintă în mişcare. Google Search a devenit din ce în ce mai bun. Şi una din noile sale funcţii a fost că Google face anumite sugestii înainte ca utilizatorul să termine ce are de scris. Se pare că aceste îmbunătăţiri au introdus câteva preferinţe noi în felul în care oamenii folosesc Google Search, iar la rândul lor acestea au făcut ca Google Flu Trends să nu mai poată funcţiona. Îmbunătăţind Google Search, am făcut să fie mai dificil de prevăzut epidemiile de gripă.

Indiferenţă

---

utilizează software special şi, în multe cazuri, şi calculatoare şi echipamente hardware special dedicate. (n. red.).

Un al cincilea scenariu este acela că orice superinteligenţă ar fi pur şi simplu indiferentă faţă de soarta noastră, la fel cum şi eu sunt indiferent faţă de soarta unor forme de viaţă mai puţin inteligente. Dacă eu construiesc o fabrică, s-ar putea să nu-mi pese neapărat că voi distruge o colonie de furnici care-mi stă în cale. Intenţia mea nu este să distrug furnicile, dar pur şi simplu ele se întâmplă să stea în locul în care trebuie să construiesc fabrica. În mod similar, o superinteligenţă nu ar fi prea preocupată de existenţa noastră. Dacă se întâmplă să fim în calea obiectivelor ei, am putea să fim eliminaţi. Superinteligenţă nu are o atitudine ostilă faţă de noi; suntem simple victime colaterale.

Pericolul unei superinteligenţe indiferente presupune că superinteligenţă nu este dependentă de umanitate. Putem distruge colonia de furnici fără nicio grijă pentru că eliminarea lor nu va provoca niciun efect advers. Dar distrugerea umanităţii ar putea avea nişte efecte adverse serioase pe care o superinteligenţă şi-ar dori să le evite. Cine oferă infrastructura pe care o foloseşte superinteligenţă? Serverele care operează în cloud? Electricitatea ce permite cloudului să funcţioneze? Internetul care face totul posibil? Dacă oamenii încă mai sunt implicaţi în vreunul dintre aceste servicii, superinteligenţa nu ar trebui să fie

indiferentă faţă de soarta noastră.

De asemenea, pericolul unei superinteligenţe indiferente presupune faptul că superinteligenţa nu ar fi paternalistă faţă de noi. De fapt, nu sunt sigur că „paternalist” ar fi adjectivul potrivit în cazul de faţă. Din cauza inteligenţei ei imense, am putea fi ca nişte copii pe care şi-ar dori să-i protejeze. Dar am fi şi părinţii lor, pe care şi-ar dori să-i protejeze drept recunoştinţă pentru că le-au adus pe această lume. Ambele sunt nişte motive bune pentru care superinteligenţa nu ar trebui să fie indiferentă faţă de existenţa noastră.

Ar trebui să ne îngrijorăm?

Aceste riscuri existenţiale depind în parte de posibilitatea ca superinteligenţa să se dezvolte foarte rapid. Dacă se întâmplă asta, vom avea foarte puţine oportunităţi, sau chiar nicio oportunitate, pentru a preveni problemele şi pentru a găsi soluţii. Dar, aşa cum am văzut până acum, există multe motive pentru care să presupunem că nu va exista o singularitate tehnologică. Dacă ar fi să se întâmple, este cel mai probabil ca superinteligenţa să se dezvolte încet, în timp ce noi muncim din greu ca să construim sisteme din ce în ce mai bune. Majoritatea colegilor mei cred că

vor mai trece multe decenii, dacă nu chiar secole, până ce va exista superinteligenţă; dacă au dreptate, ar trebui să avem suficient timp la dispoziţie ca să ne luăm măsuri de precauţie.

S-ar putea să te mai liniştească şi faptul că în ultimul deceniu s-a format o comunitate de cercetare care se preocupă de siguranţa IA. Există acum nişte grupuri de cercetare în Statele Unite, în Marea Britanie şi în alte locuri, care au primit fonduri de 10 milioane de dolari oferite de Elon Musk, pentru a căuta soluţii tehnice pentru genul de riscuri pe care tocmai le-am prezentat. Sunt destul de încrezător că, datorită acestor eforturi, IA nu va elimina umanitatea prea curând.

Totuşi, nu putem să subestimăm complet ameninţarea existenţială pe care ar putea s-o reprezinte superinteligenţa. Dar probabil ar trebui să ne concentrăm atenţia asupra unor ameninţări existenţiale mai apropiate, precum şi asupra ameninţărilor imediate pe care le implică IA şi care nu sunt existenţiale. Şi nu trebuie să mă credeţi pe cuvânt când spun că IA s-ar putea să nu fie cel mai mare risc existenţial cu care se confruntă umanitatea. Un sondaj pe 50 de laureaţi Nobel, realizat în luna septembrie a anului 2017 de *Times Higher Education*, a menţionat clima, creşterea populaţiei, războiul nuclear, bolile, egoismul, ignoranţa, terorismul, fundamentalismul şi

pe Donald Trump ca fiind nişte ameninţări mult mai mari decât IA pentru umanitate.

Pariul lui Pascal

Să presupunem pentru un moment că superinteligenţa ajunge să existe. Chiar dacă se va întâmpla peste un secol sau mai mult, oameni ca Elon Musk şi Nick Bostrom ne-au avertizat că trebuie să fim foarte îngrijoraţi. Într-adevăr, se tem că IA va ajunge să fie cea mai mare ameninţare existenţială pentru omenire, mai periculoasă decât anihilarea nucleară, încălzirea globală şi toate celelalte pericole cu care se confruntă planeta noastră în ziua de azi. Ei se tem că roboţii ne vor scăpa de sub control şi vor elimina umanitatea, în mod intenţionat sau nu.

Din punctul meu de vedere, aceste voci sunt influenţate de argumentul filosofic clasic al lui Blaise Pascal. Pascal a fost un filosof, matematician şi fizician francez care a trăit între anii 1623 şi 1662. În cartea *Cugetări*, publicată după moartea sa, Pascal a prezentat un argument despre necesitatea credinţei în Dumnezeu, ce a ajuns să fie cunoscut sub numele de „pariul lui Pascal”

Pariul lui Pascal sună aşa: Dumnezeu există sau Dumnezeu nu există, iar tu trebuie să decizi care este

adevărul. De fapt, îţi pariezi viaţa că Dumnezeu există sau nu. Acesta este pariul. Nu este opţional: dacă nu faci nimic, tot eşti de o parte anume a pariului.

Dacă tu crezi în Dumnezeu, iar El chiar există, atunci câştigi o eternitate de fericire infinită. Dacă tu crezi în Dumnezeu şi El nu există, atunci pierderea ta este finită: acţiunile din viaţa ta finită sunt guvernate de o credinţă într-un Dumnezeu care nu există. Pe de altă parte, dacă nu crezi în Dumnezeu, iar El există, atunci pierderea ta este infinită: renunţi la o viaţă eternă de fericire. Şi dacă nu crezi în Dumnezeu şi El nu există, câştigul tău a fost doar finit: ai scăpat de acţiunile inutile pe care le face un credincios.

Argumentul lui Pascal a fost unul dintre primele exemple de „teorie a deciziei", teoria logică a luării celor mai bune decizii. (Azi, IA joacă un rol central în implementarea teoriei deciziilor în calculatoare. De aceea probabil este ironic că aceleaşi ideii teoretice îi distrag pe oameni de la adevăratele ameninţări existenţiale de pe planetă.)

Teoria deciziei ne spune că a crede în Dumnezeu este superior lui a nu crede. Rişti o cantitate finită pentru a putea câştiga o cantitate infinită. Acesta este genul de pariu pe care orice bancher sau orice jucător ar trebui să-l facă fără nicio ezitare. Pariul lui Pascal vrea să te convingă logic să crezi în Dumnezeu pentru că acest

lucru maximizează recompensa pe care te poţi aştepta să o primeşti.

Dar logica şi probabilitatea nu pot decide singure dacă trebuie să credem în Dumnezeu sau nu. Oamenii preocupaţi de superinteligenţă sunt victimele unei capcane asemănătoare. Argumentul este aşa: extincţia speciei umane ar putea duce la dispariţia vieţii din univers, lăsându-l să fie un loc lipsit de viaţă. Acest lucru i-ar priva pe miliardele sau poate chiar trilioanele de oameni ce ar fi putut urma după noi de oportunitatea de a fi fericiţi. Într-adevăr, mulţi susţinători ai singularităţii tehnologice sunt transumanişti, crezând de asemenea că vom învinge moartea şi că vom deveni nemuritori. Extincţia speciei umane ar fi atunci vinovată pentru o pierdere şi mai mare de fericire.

Extincţia speciei umane este astfel o ameninţare atât de serioasă, încât trece înaintea oricăror alte griji – chiar dacă posibilitatea de a se întâmpla este extrem de mică. Este mai presus de criza financiară. Mai presus de schimbările climatice. De colapsul sistemelor noastre politice şi economice. E mai presus până şi de cele mai mari pericole pe care le-ar putea aduce Trump. Din moment ce este un risc atât de mare, suntem obligaţi să punem riscul superinteligenţei mai presus de oricare alte probleme – la fel ca în pariul lui Pascal, în care

miza fericirii eterne şi a pierderii creşte riscul lipsei de credinţă în Dumnezeu mai presus de al oricăror alte acţiuni. În ambele cazuri trebuie să ignorăm toate celelalte riscuri.

Actualizarea valorii viitoare[34]

De fapt, teoria deciziei moderne ne oferă o cale prin care putem contracara pariul lui Pascal. Soluţia este să preţuieşti prezentul, dar să pui mai puţin preţ pe viitor. Să fii recompensat astăzi valorează mai mult decât să primeşti aceeaşi recompensă mâine. Trăim în prezent şi nu în viitor. Acest lucru ne cere să fim mai cu picioarele pe pământ şi mai utilitarişti când vine vorba de recompense. Dar este probabil un punct de vedere care se potriveşte cu biologia noastră. Satisfacţia imediată este mai presus de beneficiile de pe termen lung.

O cale simplă prin care putem actualiza valoarea viitoare presupune folosirea unui factor de scontare[35]. Să presupunem că facem o reducere la fericirea din viitor de 2% pe an, rata actuală a inflaţiei în multe

---

[34] *Discounting the future*, actualizarea valorii viitoare, este un tip de preferinţă temporală şi se referă la transformarea unei valori nominale viitoare în echivalentul valoric prezent. (n. red.).

[35] Una dintre metodele cele mai puţin sofisticate de ieftinire a viitorului este considerarea unei „ferestre de timp“ fixe. Recompensele după această fereastră de timp vor fi pur şi simplu ignorate.

economii dezvoltate. Asta înseamnă că fericirea de peste un an valorează pentru tine doar 98% din fericirea de azi. Am putea să privim şi dintr-o perspectivă utilitaristă, în care însumăm fericirea populaţiei. Dacă dai la două persoane aceeaşi fericire, va fi de două ori mai valoros decât dacă-i oferi unei persoane acea fericire.

De dragul argumentului, să presupunem că populaţia globală va continua să crească în ritmul din prezent. Anul viitor, populaţia va fi cu aproximativ 1,1% mai mare decât acum. De vreme ce factorul de scontare este mai mare decât rata de creştere a populaţiei globale, fericirea populaţiei din viitor (chiar infinită) nu va avea o valoare mai mare decât a celei din prezent.

În 63 de ani, populaţia curentă a lumii se va dubla. Dar factorul de scontare al fericirii de acum împotriva celei de mâine înseamnă că fericirea unei persoane în 63 de ani valorează cât un sfert din fericirea lor de azi. Fericirea redusă a acestei populaţii dublate va face aşadar doar jumătate din fericirea totală a populaţiei globale de azi. Aşa că este mai important să facem lumea fericită astăzi decât peste 63 de ani.

În 126 de ani, populaţia actuală a globului se va cvadrupla. Dar din cauza factorului de scontare, fericirea unei persoane peste 126 de ani va face cât a

şasea parte din fericirea unei persoane din ziua de azi. Aşa că fericirea redusă a acestei populaţii cvadruple valorează cât un sfert din fericirea totală a populaţiei din prezent. Din nou, vedem că este mai important să facem lumea fericită azi decât peste 126 de ani.

Până şi un factor de scontare mic ne face să ne concentrăm eforturile pe prezent. Şi există multe riscuri foarte imediate pentru care ar trebui să ne facem griji astăzi. Aceste riscuri ne vor afecta cu mult înainte ca superinteligenţa să apuce să o facă. Într-adevăr, această carte reprezintă un avertisment cu privire la multiplele pericole reale pe care le implică IA, dintre care multe deja au început să afecteze societatea. Majoritatea implică doar IA destul de proastă pe care o avem azi, şi nu IA foarte deşteaptă pe care este posibil s-o avem peste 50 sau peste 100 de ani.

Niciodată să nu spui niciodată

Desigur, nu putem afirma cu certitudine că superinteligenţa nu va fi niciodată o ameninţare existenţială pentru societate. Istoria ne oferă o multitudine de lecţii care ne arată că niciodată nu trebuie să spui niciodată.

Colegului meu Stuart Russell îi place să ne aducă aminte că unul dintre cei mai respectaţi fizicieni

actualmente, Ernest Rutherford, a declarat că extragerea energiei dintr-un atom ar fi o „tâmpenie” şi nu va fi niciodată posibilă. A declarat asta într-un articol publicat în *The Times* pe data de 12 septembrie 1933. În ziua următoare, Leo Szilard a conceput ideea pentru reacţia nucleară în lanţ care ne-a oferit atât bomba nucleară, cât şi energia nucleară[36]. Am trecut de la „niciodată” la „înfricoşător de real” în doar 24 de ore.

Ar trebui să avem grijă când spunem „niciodată”. Arthur C. Clarke a propus trei legi care avertizează cu privire la pericolul pe care îl implică ascultarea experţilor ce prevăd ceva ce nu se va întâmpla niciodată. Merită să ne aducem aminte de ele când ne

---

[36] Potrivit istoricului Richard Rhodes, ideea unui reacţii nucleare în lanţ i-a venit lui Leo Szilard în timp ce traversa strada în Bloomsbury, Londra, a doua zi după ce Rutherford a declarat că va fi imposibil să extragi energie din atom: „Semaforul s-a schimbat în verde. Szilard a coborât de pe bordură. In timp ce traversa strada, timpul s-a deschis în faţa ochilor săi şi a văzut o cale spre viitor, moarte în lume şi pe toţi duşmanii noştri, forma lucrurilor care aveau să se petreacă" (vezi Richard Rhodes, *The Making of the Atomic Bomb*, Simon & Schuster, New York, 1986). Această reacţie nucleară în lanţ este unul dintre cele mai dramatice exemple ale puterii creşterii exponenţiale. Un neutron desparte un atom, eliminând doi neutroni. Aceşti doi neutroni împart doi noi atomi şi astfel rezultă patru neutroni. Aceştia patru creează opt, opt creează 16, 16 creează 32 şi aşa mai departe. În zece paşi, am eliberat o mie de neutroni; 20 de paşi creează un milion de neutroni, 30 un miliard, iar 40 un trilion. IA vede acum impactul unei creşteri exponenţiale similare în puterea computaţională, în date, în performanţa algoritmilor şi în finanţarea domeniului. La fel ca toate exponenţialele, o astfel de creştere nu va continua pentru totdeauna, dar, cât timp va face asta, progresele vor fi impresionante.

gândim la viitor:

PRIMA LEGE: Când un om de ştiinţă distins, dar în vârstă, declară că ceva este posibil, este aproape 100% că are dreptate. Când declară că ceva este imposibil, cel mai probabil greşeşte.

A DOUA LEGE: Singura cale prin care se pot descoperi limitele posibilului este să ne aventurăm puţin dincolo de el, în imposibil.

A TREIA LEGE: Orice tehnologie suficient de avansată nu poate fi deosebită de magie.

Aşa că permite-mi să spun că singularitatea tehnologică nu este imposibilă, şi poţi decide tu singur ce să crezi din această predicţie!

Evadarea din biologie

De fapt, singularitatea tehnologică este un fel de distragere de la ideile de bază ale superinteligenţei. Am putea ajunge la superinteligenţă fără să trecem prin niciun fel de singularitate. Computerele ar putea, spre exemplu, să devină superinteligente doar prin evitarea limitelor noastre biologice. Computerele au multe avantaje în faţa oamenilor, iar pe unele dintre ele le-am menţionat în capitolul anterior.

Evoluţia a trebuit să se obişnuiască cu diverse limite biologice. De exemplu, creierele noastre au o dimensiune limitată la aproximativ 100 de miliarde de neuroni. Şi în timp ce creierul consumă mai multă energie decât oricare alt organ din corp, este limitat la aproximativ 20 de waţi de energie. Calculatoarele nu au astfel de limite. Vrei mai multă memorie? Mai pune nişte carduri de memorie. Şi mai multă? Stochează datele în cloud. Vrei şi mai multă? Mai pune nişte amplificatoare în perete. Şi spre deosebire de oameni, calculatoarele pot lucra tot timpul. Nu au nevoie de odihnă sau de somn.

Oamenii au limite în ceea ce priveşte simţurile. Nu putem auzi la fel de bine ca un câine şi nu putem vedea la fel de bine ca un vultur. Calculatoarele pot să fie supraumane pentru că pot depăşi aceste limite. Pot recurge la noi senzori. Vedere infraroşu, precum cea pe care o au ţânţarii. Senzori ultrasonici, precum cei folosiţi de lilieci. Senzori care nu se găsesc în natură, cum ar fi GPS-ul, radarul şi lidarul. Maşinăriile pot dobândi superputeri cu uşurinţă.

De aceea maşinile autonome vor fi nişte şoferi mult mai buni decât oamenii până în anul 2062. Nu va fi aşa doar datorită faptului că sunt mai pricepute din punct de vedere tehnic, deşi cu siguranţă acest lucru va fi adevărat. În plus, vor fi capabile să calculeze distanţele

de oprire cu mai multă acurateţe decât putem noi. Nu vor fi niciodată pe viteza greşită. Nu vor încălca niciodată regulile de circulaţie. Radarul şi lidarul lor vor simţi strada mult mai bine decât ochii noştri, mai ales pe timp de vreme rea. GPS-ul nu se va rătăci niciodată. Peste 30 de ani, maşinile vor fi nişte şoferi mult mai buni decât noi, încât le vom permite şoferilor umani să conducă doar la curse sau în altfel de medii foarte controlate.

Maşinile autonome din anul 2062 vor fi nişte şoferi mai buni şi datorită puterii de concentrare supraumane. Nu vor obosi niciodată, şi nu vor fi distrase. Vor fi concentrate 100% pe condus, 100% din timp. Nu vreau o superinteligenţă care să-mi conducă maşina şi care în acelaşi timp să se gândească la cum să rezolve problema schimbărilor climatice şi la cum să aducă pacea în Orientul Mijlociu. Vreau un calculator al cărui singur scop să fie condusul.

Interfeţele creierului

Elon Musk a sugerat faptul că, pentru a putea ţine pasul cu IA, singura noastră speranţă este să creăm o interfaţă rapidă şi directă pentru creier. A finanţat o companie proaspăt apărută pentru dezvoltarea unei „plase” sau reţele neuronale ce va conecta creierele

noastre la calculatoare. Consider că argumentul lui Musk cum că acest lucru ne va permite să ţinem pasul cu calculatoarele este neconvingător, din mai multe motive.

În primul rând, deja ai o interfaţă foarte rapidă în creier: ochii. Se estimează că ochii omului au o rată de date de 10 milioane de byţi pe secundă. Aceasta este viteza portului de ethernet care îţi conectează calculatorul la internet.

În al doilea rând, rata datelor din creier nu pare să fie un factor limitator în cogniţia umană. Cel mai probabil vei putea să înveţi mai mult dacă citeşti un manual timp de două ore, decât dacă te uiţi la un film de două ore. Dar cititul timp de două ore ar însemna un aport de doar zece byţi, în comparaţie cu doi gigabyţi necesari unui film. Rata datelor din creier nu pare să ne tragă în jos. Învăţatul implică abstractizare, iar manualul este o abstractizare mult mai bună şi mai compactă.

În al treilea rând, o porţiune mare din creier este deja dedicată procesării de date de intrare. Se estimează că în jur de o treime din creier este devotată procesării văzului. Ideea că avem nişte părţi mari din creier care ar putea procesa date în plus printr-o plasă neuronală este un mit.

În al patrulea rând, dacă faci o interfaţă între creierul

tău şi un computer, cel mai probabil nu vei face decât să încetineşti computerul. Nu avem aproape niciun exemplu de oameni şi maşinării care sunt mai buni împreună decât separat. Pentru o perioadă scurtă, oamenii şi calculatoarele s-au priceput să joace şah mai bine împreună decât separat. Dar acum calculatoarele sunt atât de bune la şah, încât noi le stăm în cale. O interfaţă mai rapidă nu va face decât să ne scoată la iveală limitele. Şi asta s-ar putea să nu fie o idee prea bună, dacă te îngrijorează faptul că roboţii s-ar putea să preia controlul!

Este adevărat că nu avem o interfaţă rapidă pentru a extrage informaţii din creier. Nu avem decât o interfaţă rapidă pentru a introduce informaţii. Vorbitul şi tastatul ne permit numai să extragem un număr limitat de kilobyţi de informaţii pe minut. Dar nu par să existe semne că această viteză redusă ne trage creierul înapoi. Nu ştiu ce poţi tu, dar eu deja pot să tastez şi să vorbesc la fel de repede precum gândesc.

Să ne păstrăm avansul

Dacă o interfaţă mai rapidă nu reprezintă calea prin care oamenii pot să fie mai presus de computere, atunci ce putem face? Trebuie să apelăm la puterile noastre – la creativitate, la puterea de adaptare, la inteligenţa

emoţională şi cea socială. Dar, mai presus de toate, trebuie să ne folosim de toate acele lucruri care îi fac pe oameni speciali. Arta noastră. Iubirea noastră. Râsul. Simţul dreptăţii şi al corectitudinii. Curajul şi rezistenţa. Optimismul. Tăria de caracter. Spiritul uman. Comunitatea noastră.

Nu contează cât de inteligente vor deveni calculatoarele, vor rămâne nişte maşinării pentru totdeauna. Experienţa umană va rămâne unică. Să sperăm că *Homo digitalis* se va folosi de aceste lucruri, va externaliza acele sarcini pe care computerele le fac mai bine către computere, şi se va concentra pe experienţa umană.

0003
Sfârşitul conştiinţei

Ceea ce avem noi foarte special este conştiinţa. Din momentul în care ne trezim dimineaţa şi până în momentul în care adormim noaptea, conştiinţa este esenţială pentru experienţa noastră de viaţă. Ea reprezintă miezul a ceea ce suntem noi. Nu suntem doar inteligenţi, suntem conştienţi că suntem inteligenţi. Reflectăm asupra a ceea ce suntem. Ne facem griji. Ne aducem aminte de trecut. Planificăm viitorul.

Uneori, această conştientizare constantă devine atât de chinuitoare, încât căutăm să trăim experienţe care să ne distragă mintea conştientă. Să trăim pur şi simplu clipa. Meditătm. Cântăm. Alergăm maratoane. Bem alcool. Sărim de pe o stâncă.

Ce este conştiinţa? În ce fel este legată de inteligenţă? Vor ajunge calculatoarele să fie conştiente vreodată? Până în anul 2062, oare IA ne va permite să ne încărcăm conştiinţa în cloud? Oare *Homo digitalis* va avea o conştiinţă care va fi şi biologică, şi digitală?

Problema dificilă

Filosoful australian David Chalmers a numit

conştiinţa „problema dificilă”. Unii au susţinut chiar că problema este mult prea grea pentru minţile noastre sau că nici măcar nu ţine de domeniul cercetărilor ştiinţifice. Chalmers este de părere că şi conştiinţa va fi la un moment dat înţeleasă din punct de vedere ştiinţific, dar în anul 1995 a afirmat că deocamdată ne lipseşte ceva important: „Pentru a reprezenta experienţa conştientă, avem nevoie de un ingredient în plus în explicaţie. Acest lucru constituie o provocare pentru cei care se gândesc serios la problema conştiinţei. Care este acel ingredient în plus, şi de ce reprezintă experienţa conştientă?”[37]

Câţiva ani mai târziu, a observat:

Conştiinţa este cea mai înşelătoare problemă a ştiinţei cu privire la minte. Nu există nimic altceva ce cunoaştem mai intim decât experienţa conştientă, dar în acelaşi timp nu există nimic mai complicat de explicat. Tot felul de fenomene mentale s-au pretat la investigare ştiinţifică în ultimii ani, dar conştiinţa a rezistat cu încăpăţânare. Mulţi au încercat să o explice, dar explicaţiile par să fie nesatisfăcătoare.[38]

---

[37] David Chalmers „Facing Up to the Problem of Consciousness”, Journal of Consciousness Studies, vol. 2, nr. 3, 1995, pp. 200-219.

[38] David Chalmers „The Singularity: A Philosophical Analysis“, Journal of Consciousness Studies, vol. 17, nr. 9-10, 2010, pp. 7-65.

Nu avem niciun instrument pentru măsurarea conştiinţei. Din câte ne putem da seama, nu există nicio parte anume din creier responsabilă pentru conştiinţă. Fiecare dintre noi ştie că este conştient – şi având în vedere că suntem înrudiţi biologic, cei mai mulţi dintre noi suntem pregătiţi să acceptăm faptul că şi alţi oameni sunt conştienţi.

Atribuim niveluri limitate de conştiinţă şi anumitor animale. Câinii au un anumit nivel de conştiinţă, la fel şi pisicile. Dar nu mulţi dintre noi cred că furnicile au ceva de genul conştiinţei. Dacă te uiţi la maşinăriile pe care le construim în prezent, atunci poţi să fii destul de sigur de faptul că nu au absolut nicio conştiinţă.

Alphago nu se va trezi într-o dimineaţă să spună: „Ştii ce? Voi, oamenii, nu vă pricepeţi deloc la Go. Mai bine fac eu bani jucând poker online". În realitate, Alphago nici măcar nu ştie că joacă Go. Va face mereu acelaşi lucru: va maximiza probabilitatea să câştige jocul respectiv. Şi cu siguranţă nu se va trezi şi nu se va gândi: „De fapt, m-am săturat să joc jocuri. O să preiau controlul asupra planetei". Alphago nu are alte scopuri decât să maximizeze probabilitatea câştigului. Nu are nicio dorinţă. Este şi va fi pentru totdeauna doar un program care joacă Go. Nu se va întrista când va pierde, şi nu se va bucura când va câştiga.

Dar nu putem fi siguri că această situaţie nu se va schimba. Poate că pe viitor vom construi calculatoare care vor fi conştiente. Pentru a se comporta etic, ar putea fi foarte important ca acele calculatoare să reflecteze asupra deciziilor luate. Pentru a acţiona într-o lume nesigură, în continuă schimbare, va fi necesar să construim maşinării cu scopuri foarte deschise şi cu abilitatea de a reflecta asupra felului în care vor atinge aceste obiective şi cum le vor adapta. Aceştia ar putea fi paşii pe drumul spre o maşinărie conştientă.

Conştiinţa este o parte atât de importantă a existenţei noastre, încât trebuie să ne întrebăm dacă ne-a oferit un avantaj evolutiv esenţial. Societăţile noastre complexe funcţionează în parte deoarece suntem conştienţi de cum ar putea gândi ceilalţi. Şi dacă conştiinţa a fost un avantaj evolutiv atât de mare pentru oameni, s-ar putea să fie util să le oferim roboţilor acest avantaj.

Maşinării conştiente

Conştiinţa la maşinării ar putea apărea pe trei căi: ar putea fi programată, ar putea apărea în urma complexităţii sau ar putea fi învăţată.

Prima rută pare una dificilă. Cum programăm ceva ce înţelegem atât de puţin? S-ar putea să fie de ajuns să

punem un strat executiv pe deasupra automatului, care monitorizează acţiunile şi motivele sale. Sau e posibil să trebuiască să aşteptăm până ce înţelegem conştiinţa mai bine, pentru a începe să o programăm.

Ca alternativă, conştiinţa s-ar putea să nu necesite să fie programată în mod explicit. Ar putea să fie un simplu fenomen emergent. Avem o mulţime de exemple de fenomene emergente în sistemele noastre complexe. Viaţa, spre exemplu, a apărut ca fenomen emergent din universul nostru complex. Conştiinţa ar putea şi ea în mod similar să emeargă dintr-o maşină suficient de complexă. Asta se întâmplă în natură, conştiinţa este legată de creierele mai mari şi mai complexe. Ştim atât de puţin despre conştiinţă, încât trebuie să luăm în calcul posibilitatea că ar putea să se producă de la sine într-un sistem IA suficient de complex.

A treia opţiune – aceea că automatele învaţă să fie conştiente – este şi ea rezonabilă. O mare parte din conştiinţa *noastră* pare să fie învăţată. Atunci când ne naştem, avem o conştiinţă limitată. Ne bucurăm că ne descoperim degetele de la picioare. Ne ia mai multe luni, dacă nu chiar un an sau mai mult, ca să ne dăm seama că imaginea dintr-o oglindă este, de fapt, a noastră. Dacă învăţăm să fim conştienţi de noi înşine, nu ar putea şi o maşină să facă la fel?

Totuşi, conştiinţa ar putea fi ceva ce nu poate fi simulat într-o maşinărie. Ar putea fi proprietatea unui anumit tip de materie. O analogie potrivită este cea cu vremea. Poţi simula o furtună în calculator, dar înăuntrul acestuia nu va fi niciodată ud. În mod similar, conştiinţa ar putea să apară doar prin materie aranjată corect. Iar siliciul ar putea să nu fie materia potrivită pentru a se ajunge la conştiinţă.

Inteligenţă de zombi

Când le vorbesc oamenilor despre inteligenţa artificială, se concentrează adesea pe primul cuvânt, „inteligenţă". La urma urmei, inteligenţa este cea care ne face speciali, şi asta este ceea ce îşi propune IA să realizeze. Dar eu vreau să le amintesc oamenilor să se gândească şi la cuvântul „artificial". În anul 2062 s-ar putea să fi construit un alt fel – foarte artificial – de inteligenţă, diferită de cea naturală pe care o avem noi. S-ar putea, spre exemplu, să nu fie o inteligenţă conştientă.

Zborul este o analogie bună. Am reuşit cu succes să construim zboruri artificiale. Am construit avioane care pot zbura mai rapid decât viteza sunetului, pot traversa oceane în câteva ore şi pot transporta tone de încărcătură. Dacă am fi încercat să recreăm zborul

natural, îmi închipui că încă am fi la capătul pistei încercând să dăm din aripi. Am abordat problema zborului dintr-o perspectivă diferită de cea a naturii – cu o aripă fixă şi un motor puternic. Atât zborul natural, cât şi cel artificial depind de aceeaşi teorie a aerodinamicii, dar sunt nişte soluţii diferite cu privire la provocarea zborului. Iar natura nu găseşte neapărat cea mai uşoară sau cea mai bună soluţie[39].

În mod similar, în comparaţie cu inteligenţa naturală, cea artificială reprezintă o soluţie complet diferită pentru problema noastră. Este posibil să fie o formă inconştientă de inteligenţă. David Chalmers a numit-o „inteligenţă zombi". Am putea construi IA care să fie inteligente, chiar superinteligente, dar care să nu posede nicio formă de conştiinţă. Pentru noi ar fi cu siguranţă inteligenţă *artificială.* Inteligenţa zombi ar demonstra că inteligenţa şi conştiinţa sunt oarecum

---

[39] Pe data de 9 octombrie 1903, *New York Times* a publicat un editorial care s-a numit „Maşinile zburătoare care nu zboară" Desfiinţa ideea cum că oamenii vor crea maşini zburătoare în viitorul apropiat. „Eşecul ridicol al încercării de navigaţie aeriană a maşinilor zburătoare Langley nu a fost ceva neaşteptat pentru nimeni, excepţie făcând probabil doar distinsul Secretar al Institutului Smithsonian, cel care le-a creat, şi asistenţii săi... Putem asuma faptul că maşinile zburătoare care chiar vor zbura vor evolua probabil din eforturile combinate şi continue ale matematicienilor şi ale mecanicilor peste un milion până la zece milioane de ani. " In realitate, nu a durat un milion de ani. După doar 69 de zile, fraţii Wright au demonstrat zborul susţinut al unei aeronave mai grele decât aerul pe o plajă unde bătea briza, în apropiere de Kitty Hawk, Carolina de Nord. Omenirea a găsit o modalitate de a zbura cu mult diferită faţă de natură.

nişte fenomene separate. Că poate exista inteligenţă fără conştiinţă. Dar dacă nu poţi să dezvolţi conştiinţă fără inteligenţă, ceea ce pare şi mai probabil, inteligenţa zombi nu ar demonstra faptul că cele două fenomene sunt complet separabile. Una o implică pe cealaltă, dar nu şi viceversa.

Inteligenţa zombi ar fi un noroc pentru omenire, din punctul de vedere al eticii. Dacă IA ar avea doar inteligenţă zombi, atunci s-ar putea să nu trebuiască să ne îngrijorăm din cauza felului în care tratăm maşinile. Le-am putea pune să îndeplinească cele mai obositoare şi cele mai repetitive sarcini. Am putea să le oprim fără să ne gândim că ar suferi. Pe de altă parte, dacă IA nu va fi o inteligenţă zombi, s-ar putea să fim nevoiţi să luăm nişte decizii dificile din punct de vedere etic. Dacă maşinile devin conştiente, vor avea drepturi? Există campanii care luptă pentru acordarea de drepturi primatelor mari, cum ar fi cimpanzeii şi gorilele. Roboţii inteligenţi ar avea şi ei aceleaşi drepturi dacă ar deveni conştienţi? Ar trebui să mai avem atunci posibilitatea de a-i opri?

Extraordinara caracatiţă

Oamenii nu sunt singura specie conştientă de pe planetă. Natura a descoperit cel puţin două rute

diferite în ceea ce priveşte inteligenţa şi conştiinţa. Cefalopodele – în special cele din subclasa Coleoidea, care includ nevertebrate cum ar fi caracatiţa – au un sistem nervos extrem de diferit faţă de vertebrate. Se crede că cefalopodele sunt cele mai inteligente nevertebrate.

Poate că nu sună prea impresionant, întrucât un număr de nevertebrate, cum ar fi moluştele, nu au creier. Dar inteligenţa cefalopodelor este uluitoare. Ele folosesc instrumente. Cooperează şi comunică unele cu altele atunci când vânează împreună. Pot să deschidă borcane cu capac înfiletat. Cefalopodele par chiar să fie capabile să recunoască şi să-şi aducă aminte de oameni. Se spune că Otto, o caracatiţă faimoasă de la acvariul Sea Star din Coburg, Germania, putea să închidă un reflector enervant de 2000 de waţi pe timpul nopţii, stropindu-i cu apă. Există multe astfel de anecdote care demonstrează abilitatea lor de a rezolva probleme, inclusiv tendinţa de a evada din bazine.

Datorită inteligenţei lor, cefalopodele sunt protejate şi nu se fac teste pe ele în mai multe ţări. În Marea Britanie, caracatiţa comună a fost singura nevertebrată protejată de *Animals (Scientific Procedures) Act* din anul 1986. Iar în Europa cefalopodele sunt singurele nevertebrate protejate de directiva europeană din 2010 cu privire la protecţia animalelor în cadrul

experimentelor ştiinţifice.

Cefalopodele au apărut cu aproximativ 600 de milioane de ani în urmă, cu mult înainte ca dinozaurii să domine planeta. La acea vreme, cele mai complexe animale nu aveau decât câţiva neuroni. Cefalopodele şi-au dezvoltat inteligenţa diferit faţă de noi. Trei cincimi din neuronii unei caracatiţe se află în cele opt tentacule ale sale. Fiecare tentacul simte şi poate gândi independent. Într-un fel e ca şi cum fiecare tentacul ar avea propriul creier. E greu să îţi imaginezi cum ar fi să fii o caracatiţă, iar picioarele tale să facă ce vor – probabil e o senzaţie ca atunci când eşti beat.

Această formă de inteligenţă exotică şi ciudată ne aduce aminte de faptul că într-o zi s-ar putea ca IA să aibă o inteligenţă diferită faţă de a noastră. Chiar dacă vom construi o IA care să fie conştientă, tot s-ar putea să aibă o inteligenţă foarte *artificială* în comparaţie cu a noastră. Şi aceasta ar fi încă o situaţie norocoasă în care am putea fi scutiţi de griji etice.

Vacile, porcii şi multe alte animale pe care le-am domesticit au un anumit nivel de inteligenţă şi de conştiinţă. Dar pentru că sunt suficient de „diferite” faţă de noi, părem pregătiţi să trecem peste faptul că le tratăm destul de prost. În mod similar, am putea dori să tratăm o inteligenţă *artificială* în moduri care ar fi inacceptabile dacă ar fi om.

Probleme dureroase

Maşinăriile din ziua de azi nu simt durere. Poate că acest lucru nu este surprinzător, de vreme ce nu sunt conştiente, iar durerea pare strâns legată de conştiinţă. Drepturile umane pot fi văzute ca fiind un răspuns moral la prevenirea durerii şi a suferinţei. Aşa că, fără durere, poate că automatele nu au niciun drept?

Dintr-o perspectivă tehnologică, ar putea să fie util pentru automatele din anul 2062 să simtă durerea. Dacă aceasta ar trebui să fie o durere adevărată, ca a noastră, sau una artificială, rămâne de dezbătut. Durerea are o istorie evolutivă foarte lungă, şi se pare că le-a priit atât oamenilor, cât şi animalelor. Durerea îi poate ajuta pe roboţi să înveţe mai rapid din propriile greşeli. Ne luăm mâna din foc imediat nu pentru că ne dăm seama cât de mult ne-ar dăuna, ci datorită intensităţii durerii. Roboţii ar putea să beneficieze şi ei de pe urma unor mecanisme de feedback similare.

Dar a le înzestra pe automate cu simţul durerii s-ar putea întoarce împotriva noastră. Atunci ar fi capabile să sufere, şi astfel s-ar cuveni să aibă drepturi care să prevină această suferinţă. Acest aspect le-ar limita utilitatea extrem de mult pentru că nu am mai putea să le forţăm să ne facă treburile murdare, cele mai grele şi

mai periculoase. Şi dacă le oferim roboţilor durere, poate că ar trebui să le oferim şi simţul fricii, care adesea este legat de durere şi astfel previne rănirea. Şi de ce să ne oprim acolo? Nu ar fi folositor ca roboţii să aibă şi alte emoţii umane, cum ar fi dorinţa, fericirea, interesul, surprinderea, mirarea şi chiar tristeţea?

Oamenii au vieţi emoţionale complexe. Într-adevăr, emoţiile noastre şi inteligenţa par să fie conectate foarte strâns. O mare parte din comportamentul nostru este dictată de emoţii. Acestea par să ofere un avantaj evolutiv semnificativ: surprinderea ne ajută să identificăm ce e nou şi potenţial periculos; curiozitatea ne împinge să descoperim şi să controlăm lumea. Nu ne-am dori ca IA să aibă şi ea această abilitate?

În prezent, computerele nu au nicio emoţie. Dar au fost făcuţi primii paşi pentru a fi se oferi o înţelegere rudimentară cu privire la emoţiile umane. Pot, spre exemplu, să decidă dacă un e-mail „sună" supărat. Acest lucru le va ajuta să interacţioneze cu noi în cadrul conversaţiei pe termen lung pe care o vom avea cu roboţii.

Filmul *Ea* (*Her,* 2013) a prezentat lucrurile exact aşa cum trebuie. Inteligenţa artificială este sistemul de operare al anului 2062. Vom începe să interacţionăm din ce în ce mai mult cu toate dispozitivele conectate la vieţile noastre vorbind cu ele. Nu vor exista tastaturi,

vom realiza imputul şi outputul de informaţie doar vocal şi auditiv. Şi această conversaţie te va urma în orice cameră, în maşină, la birou şi apoi înapoi acasă.

Pentru a face aceste conversaţii şi mai antrenante, inginerii vor fi tentaţi să le ofere maşinilor propriile lor vieţi emoţionale. Din nou, rămâne de văzut dacă vor fi emoţii reale sau simulate. Dar pare extrem de posibil ca automatele inteligente care vor interacţiona cu noi în anul 2062 să fie emoţionale.

Marvin, androidul paranoic din *Ghidul autostopistului galactic* al lui Douglas Adams, a exprimat bine acest lucru: „Şi mai e desigur şi durerea asta groaznică pe care o simt în toate diodele din partea stângă a mâinii...” Pe lângă faptul că avea diode dureroase, Marvin a devenit de asemenea faimos pentru că era un robot depresiv:

ARTHUR DENT: (Pământul) a fost un loc frumos.

MARVIN: Avea oceane?

ARTHUR DENT: A, da; oceane albastre, extrem de întinse.

MARVIN: Nu-mi plac oceanele.

Aceste trăsături umane i-au înduioşat pe mii de fani ai *Ghidului autostopistului.* Ne aşteptăm ca până în anul 2062 să fim şi noi mişcaţi şi emoţionaţi de multe alte

maşinării.

Problema cu liberul arbitru

Chiar dacă se dovedeşte că nu este imposibilă conştiinţa pentru IA, trebuie să ne gândim la încă o problemă controversată şi legată de filosofie. Problema liberului arbitru. Inteligenţa artificială nu pune într-o lumină bună această provocare veche, deoarece computerele sunt maşini deterministe. Ele urmează instrucţiuni precise. Dispozitivele digitale, cum ar fi calculatoarele, sunt poate cele mai deterministe invenţii ale noastre.

Într-un dispozitiv analog, cum ar fi un termostat, există toleranţe şi incertitudini care înseamnă că se poate comporta în feluri neaşteptate. Până şi sistemele simple pot să aibă nişte bucle de feedback haotice care împiedică predicţiile felului în care vor acţiona. Într-un dispozitiv digital nu există astfel de ambiguităţi. Starea este ori 0, ori 1. Nu există nicio incertitudine sau hazard. Calculatoarele sunt interpretarea fizică a logicii matematice, aşa că sunt la fel de exacte şi de lipsite de ambiguitate ca logica. Dacă aduni 1 cu 1, mereu îţi va da 2. Dacă ceri rădăcina pătrată a numărului 9, vei obţine mereu 3.

Atunci cum poate o IA, un program care rulează pe

un calculator determinist, să aibă vreun fel de liber arbitru? Liberul arbitru este unul dintre acele subiecte pe care oamenii de ştiinţă au grijă să le evite. Ca oameni de ştiinţă, cei mai mulţi dintre noi ne punem noaptea în pat fiind conştienţi de faptul că nu prea avem cum să-l explicăm. Dezvoltăm ecuaţii matematice care guvernează precis momentul în care acţionează un neuron. Propunem căi chimice care au rezultate precise. Discutăm transferul de ioni care este condus de ecuaţii diferenţiale exacte. Unde să existe loc în aceste modele ale creierului pentru liberul arbitru?

Liberul arbitru nu este o problemă ce ţine doar de IA. Este un lucru pe care ştiinţa nu poate să-l explice în creierele noastre biologice. Modelele ştiinţifice de creier lasă nişte spaţii acolo unde liberul arbitru s-ar putea ascunde. Un astfel de spaţiu este în zona cuantică.

Roger Penrose, unul dintre cei mai vehemenţi critici la adresa posibilităţii inteligenţei artificiale, susţine că creierul nu este algoritmic, şi nu poate fi modelat de un calculator digital convenţional. El bănuieşte că mecanica cuantică joacă un rol esenţial în creierul uman[40]. Juriul ştiinţific încă nu se poate decide dacă mecanica cuantică joacă într-adevăr acest rol esenţial.

---

[40] Pentru mai multe discuţii legate de posibilele efecte cuantice din creierul uman, vezi Roger Penrose, *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics,* Editura Oxford University, New York, 1989.

Dar chiar dacă ar fi aşa, nimic nu poate opri un computer digital să simuleze mecanica cuantică, ceea ce ne face să ne întrebăm dacă Penrose a găsit cumva un loc în care să se ascundă o fantomă.

Un alt loc în care s-ar putea ascunde o fantomă este complexitatea lumii în care interacţionăm. Liberul arbitru ar putea să fie doar ceva ce reiese din interacţiunea dintre oameni şi mediu. După cum a afirmat Daniel Dennett, liberul arbitru ar putea fi o simplă iluzie care reiese din această complexitate.

Adesea uităm să ne aducem aminte că felul în care luăm decizii este modelat de o lume complexă. Ceea ce pentru noi pare liber arbitru nu este nimic mai mult decât un sistem mare, foarte complex, în care suntem nişte jucători minusculi. Indiferent dacă vom afla că liberul arbitru este real sau o iluzie, IA ne poate ajuta să rezolvăm misterul.

Roboţi etici

Dacă maşinăriile inteligente ajung să aibă liber arbitru, atunci va trebui să ne preocupe valorile lor. Se vor folosi de acest liber arbitru pentru a face lucruri rele? Chiar şi fără liber arbitru, vrem ca maşinăriile inteligente să se comporte etic.

Problema nu o reprezintă doar maşinile *deştepte.* Am

vrea ca un IA neinteligent să se poarte, de asemenea, etic. Într-adevăr, avem această problemă în prezent în cazul maşinilor autonome. Acestea nu sunt prea inteligente, încă. Dar deja le permitem să ia decizii de viaţă şi de moarte.

Când vine vorba de comportamentul etic, o maşinărie conştientă va avea un avantaj în faţa uneia inconştiente. Poate să reflecteze asupra acţiunilor sale, atât cele din trecut şi prezent, cât şi cele din viitor. Poate să reflecteze asupra felului în care ceilalţi vor fi afectaţi de aceste acţiuni. Este greu de imaginat că poate fi construit un robot care să se comporte etic fără ca acesta să aibă un nivel executiv care să funcţioneze, într-un anumit fel, ca un eu uman conştient.

Ce ar trebui să fie inclus în acest nivel executiv care va conduce felul în care acţionează un robot? Ar trebui poate să programăm ceva de genul celor trei legi de robotică ale lui Asimov? Isaac Asimov a propus nişte legi ce au devenit faimoase, încă din anul 1942.

PRIMA LEGE: Un robot nu poate să rănească o fiinţă umană sau, prin inacţiune, să permită ca un om să fie rănit.

A DOUA LEGE: Un robot trebuie să respecte ordinele care îi sunt date de către oameni, cu excepţia celor care ar intra în conflict cu Prima Lege.

A TREIA LEGE: Un robot trebuie să-şi protejeze propria existenţă, atât timp cât această protecţie nu intră în conflict cu Prima sau A Doua Lege.

Potrivit lui Asimov, aceste legi proveneau din *Manualul de robotică*, a cincizeci şi opta ediţie, publicată în anul 2058. Sunt sigur că Asimov va avea dreptate şi că până în anul 2058 vor exista manuale de instrucţiuni etice bine definite pentru construirea roboţilor. Dar ceea ce povestea lui Asimov demonstrează este faptul că cele trei legi ale sale au nişte limite. Ce se întâmplă dacă a întreprinde o acţiune va răni o persoană, dar a nu face nimic va răni o alta? Ce se întâmplă dacă doi oameni dau ordine contradictorii?

Un experiment mental clasic din filosofie este un bun exemplu de dilemă morală care constituie o provocare pentru legile lui Asimov. Să presupunem că stai în maşina ta Tesla, ce conduce singură pe o stradă. Deodată, doi copii mici aleargă pe stradă după o minge. Ce va decide Tesla, va da peste cei doi copii, sau va vira şi se va izbi de o maşină parcată?[41] Computerul are la dispoziţie câteva milisecunde ca să decidă ce acţiune să aleagă, iar ambele ar avea ca rezultat răni sau chiar moarte. Nu mai este suficient timp ca să poţi

[41] Vezi şi „problema trenului" (*the trolley problem*) descrisă la p. 164. (n. red.).

prelua tu controlul. Dacă nu face nimic, cel mai probabil copiii vor muri. Dar dacă acţionează evaziv şi se loveşte de o maşină parcată, cel mai probabil tu vei fi cel care va muri. Legile lui Asimov nu o ajută pe Tesla să decidă ce să facă.

Limitele legilor lui Asimov sunt evidenţiate de introducerea ulterioară de către el a unei a patra legi:

LEGEA ZERO: Un robot nu poate să rănească umanitatea sau, prin inacţiune, să permită umanităţii să fie rănită.

Această lege are numărul zero întrucât le precedă pe cele trei prezentate anterior. Principiul face referinţă la faptul că sunt situaţii în care, dacă un robot răneşte un om, s-ar putea ca asta să fie cea mai bună soluţie. Dar tot nu ne lămureşte ce ar trebui să facem mai exact în situaţia cu maşina Tesla.

Principiul zero al lui Asimov creează de fapt un set nou de probleme. Cum poate un robot să decidă ce anume va răni umanitatea? Ce înseamnă de fapt rănirea umanităţii? Cum cântărim, spre exemplu, binele oamenilor care sunt în viaţă în raport cu al celor care încă nu s-au născut?

Bănuiesc faptul că nu vom reuşi să programăm manual valorile etice ale maşinăriilor noastre

inteligente. În schimb, poate vom decide să le permitem să înveţe să se comporte etic. Dar şi asta pune provocări. Adesea, noi, oamenii, acţionăm în feluri care nu sunt în concordanţă cu standardele etice pe care ni le-am stabilit chiar noi. Aşadar, automatele vor întâmpina dificultăţi în învăţarea valorilor etice prin observarea comportamentului uman.

De fapt, probabil că roboţii vor avea standarde etice mai înalte decât noi. Ei nu posedă nicio slăbiciune umană. O maşină autonomă poate fi programată să nu meargă niciodată cu viteză mare. O dronă autonomă şi înarmată poate fi programată să urmeze întotdeauna legile umanitare internaţionale. Iar un avocat robot poate fi programat să nu ascundă niciodată dovezile descoperite. Este o întrebare deschisă şi interesantă: cum vor putea fi învăţate astfel de valori superetice?

Sfârşitul morţii

Conştiinţa noastră este strâns legată de faptul că suntem în viaţă. Şi din ce ne poate spune ştiinţa, conştiinţa pare că sfârşeşte odată cu moartea. Şi astfel ajungem la subiectul morţii. Şi ca să ajungem la moarte, vom trece prin transumanism.

Inteligenţa artificială din ziua de azi a ajuns să fie destul de apreciată. Cu 20 de ani în urmă, când le

spuneam oamenilor că lucrez în domeniul inteligenţei artificiale, râdeau de mine şi spuneau cât de proaste sunt calculatoarele. Dar acum, de vreme ce IA începe să se infiltreze în vieţile noastre şi chiar să îmbunătăţească unele părţi ale lor, oamenii spun adesea: „Lucrezi în IA – ce mişto!” Inteligenţa artificială generală, pe de altă parte, rămâne încă în zona de avangardă. Şi dacă începi să explorezi acea avangardă, s-ar putea să ajungi la transumanism. Aici vei găsi oameni care văd IA ea pe o modalitate prin care poţi păcăli moartea.

E interesant că mulţi dintre cei care ne avertizează cu privire la riscurile pe care le presupune IA, oameni ca Nick Bostrom, sunt transumanişti. Bineînţeles, acest lucru nu ar trebui să ne surprindă prea mult. Dacă ai de gând să trăieşti pentru totdeauna, atunci ai avea foarte mult de pierdut personal dacă IA ar ajunge să elimine specia umană. Aşadar, este amuzant într-un mod ironic să constaţi că unii transumanişti îşi doresc ca IA să fie posibilă, întrucât îi va ajuta să îşi îndeplinească visurile transumaniste, dar, în acelaşi timp, se tem de ce ar putea face IA.

Cum ar putea pune IA capăt morţii? Corpurile noastre biologice sunt nişte recipiente imperfecte. Virusurile şi infecţiile bacteriene ne invadează sistemul de apărare. Sistemul nostru imunitar poate ceda şi cancerul ne năpădeşte şi ne omoară. Mecanismele

noastre de reparare nu mai funcţionează după o perioadă, aşa că îmbătrânim şi ne degradăm. După vârsta de 70 de ani, majoritatea suntem pe drumul spre morgă. Siliciul ar fi un adăpost cu mult superior pentru inteligenţa noastră. Este mai rapid, mai încăpător şi nu se degradează. Şi de vreme ce informaţia digitală poate fi copiată fără nicio eroare, ne putem oricând muta într-o locuinţă nouă perfectă.

Aşa că, în anul 2062, pur şi simplu ne vom încărca minţile în cloud şi vom duce o existenţă virtuală? Având în vedere caracteristicile superioare ale unui substrat digital, oare *Homo digitalis* va fi complet digital? Sau va rămâne jumătate biologic şi jumătate digital? Răspunsurile la aceste întrebări depind într-o oarecare măsură de natura conştiinţei. Dacă aceasta poate fi doar biologică, atunci când ne vom muta în cloud vom lăsa o parte importantă în urmă. Sinele nostru zombi ar putea să iasă la iveală. Va şti tot ce ştim noi. Va spune toate lucrurile pe care le spunem noi. Dar nu vom fi *noi.* Pe de altă parte, dacă IA poate să fie conştientă, atunci probabil că sinele nostru încărcat în reţea va moşteni o conştiinţă asemănătoare cu cea a sinelui nostru biologic. Dar aceasta este o posibilitate mai neliniştitoare.

Din fericire, nişte limitări tehnologice ne-ar putea scuti de dilemele etice pe care le-ar pune încărcarea

conţinutului creierului într-un computer. Ar putea să fie pur şi simplu imposibil sau ineficient să citim conţinutul creierelor noastre biologice cu destulă acurateţe, astfel încât să creăm o copie digitală. Creierul uman, cu miliardele lui de neuroni şi trilioanele de sinapse, este cel mai complex sistem din univers pe care îl cunoaştem, cu mai multe ordine de mărime. S-ar putea să nu fie fezabil din perspectivă tehnică să „citim" toate aceste lucruri. Sau poate va fi posibil, dar numai într-un mod distructiv. Adică am reuşi să citim conţinutul creierului biologic, dar făcând asta am distruge originalul. Astfel, am putea reuşi să creăm o copie artificială, dar rămânând fără un creier biologic funcţional. Această variantă ne-ar scuti de nişte probleme morale şi etice dificile, legate de identitate şi sine.

Viaţa virtuală

Mă aştept să ne păstrăm forma biologică, dar o vom augmenta prin asistenţi digitali care ne vor oferi ajutor constant. Vom avea şi avataruri digitale, care vor suna şi se vor comporta ca noi în lumea digitală. Şi până în anul 2062, ne vom cufunda în lumea virtuală şi în realitatea augmentată atât de mult, încât ne va fi imposibil să o deosebim de lumea reală. Chiar dacă

trupul nostru fizic se va afla în lumea reală, creierele noastre vor crede că se află în lumea digitală.

Aceste lumi virtuale vor fi seducătoare. Cu toţii vom putea fi bogaţi şi faimoşi în astfel de lumi. Vom putea să fim frumoşi şi deştepţi. Vom avea succes. Nu va fi nimic care să ne stea în cale.

Lumea reală va fi indubitabil mai puţin plăcută. Nu va fi surprinzător că unii dintre noi vor deveni dependenţi de evadarea din realitate pe această cale, de petrecerea fiecărui moment, cât suntem treji, în lumile virtuale. Deja vedem cum astfel de tendinţe se petrec în lumea dezvoltată. În anii 2000, bărbaţii din Statele Unite cu vârste cuprinse între 21 şi 30 de ani fără diplomă de licenţă au avut cu 20% mai puţine ore lucrate. În acelaşi timp, mulţi dintre aceşti bărbaţi îşi petrec o mare parte din timp jucând jocuri video. Într-adevăr, mulţi petrec mai mult timp jucând jocuri decât socializând cu prietenii, practicând sporturi sau alte hobby-uri. Acest grup constituie 40% din populaţia totală de jucători cu vârste cuprinse între 21 şi 55 de ani, în ciuda faptului că reprezintă 10% din acel segment de vârstă din populaţie. Pentru acest grup, lumea virtuală începe să ofere din ce în ce mai mult un loc virtual de evadare din lumea reală.

Un aspect tulburător al lumilor virtuale va fi că unii oameni vor locui în ele pentru a se purta în feluri care

nu sunt acceptabile în lumea reală. Societatea va trebui să decidă dacă acţiunile care sunt ilegale în lumea reală vor fi făcute ilegale sau imposibile în cea virtuală. Sau societatea va decide că astfel de lumi virtuale sunt o necesară supapă de siguranţă. Va fi o decizie dificilă, şi mă aştept ca opiniile să oscileze.

Sfârşitul biologiei

O altă cale prin care am putea pune capăt morţii este prin înfrângerea biologiei. IA ne va ajuta să înţelegem mai multe despre felul în care funcţionează corpul uman, şi despre cum să oprim sau chiar să inversăm procesul îmbătrânirii. Poate că vom reuşi să vindecăm toate bolile care ne afectează, inclusiv boala care ia viaţa celor mai mulţi din lumea occidentală: bătrâneţea.

De ce trăim 70 de ani, şi nu 700 de ani? Şi dacă putem ajunge la 700 de ani, de ce să ne oprim acolo? În prima jumătate a vieţii, corpurile noastre se descurcă foarte bine să se repare singure. Am putea păcăli corpul să facă asta pentru o perioadă mai lungă? În combinaţie cu tehnologii noi de editare a genelor cum ar fi CRISPR[42], IA ne-ar putea ajuta să ne apropiem de

[42] CRISPR (pronunţat „crispăr") vine de la Clustered Regularly Interspaced Short Palindromic Repeats. Reprezintă baza unei tehnologii noi puternice de editare a genelor. Revista *Science* a numit-o descoperirea anului 2015. Cu

nemurire.

O astfel de schimbare pentru existenţa umană ar duce la modificări extreme ale societăţii. Dacă această „nemurire” va fi accesibilă doar celor bogaţi, societatea noastră va deveni şi mai divizată, se va împărţi în bogaţii nemuritori şi săracii muritori. Oamenii bogaţi deja trăiesc mai mult decât cei săraci, dar o diferenţă atât de mare ar provoca un mare conflict social.

Pe de altă parte, dacă nemurirea de genul acesta ar fi disponibilă pentru toată lumea, va trebui să regândim cum va funcţiona societatea. Va fi necesar să introducem eutanasia voluntară pentru a face loc noilor generaţii? Va trebui să avem o lege prin care să interzicem oamenilor să se reproducă şi poate o loterie prin care se va decide când poţi face copii pentru a-i înlocui pe oamenii care mor accidental sau intenţionat? Cum vom reinventa ciclul copilăriei, al muncii şi al pensionării pentru vieţi care ar putea dura sute sau chiar mii de ani?

În cazul unei vieţi fără sfârşit va trebui să ne gândim din nou care este înţelesul existenţei noastre. Viaţa este scurtă, şi asta face parte din frumuseţea ei. Trebuie să trăim viaţa la maximum pentru că nu ştie nimeni când se va termina. Şi cu toţii avem în comun aceeaşi soartă

---

siguranţă comitetul de la premiile Nobel îi va recunoaşte potenţialul în viitorul apropiat.

finală. Dacă digitalis va transcende asta, va trebui să găsim noi lucruri care să dea vieţii sens.

0004
Sfârşitul muncii

Să nu uităm faptul că „muncă” este singurul cuvânt cu adevărat obscen. Una dintre cele mai vizibile consecinţe ale inteligenţei artificiale din anul 2062 va fi că oamenii vor munci mult mai puţin. Viitorul ne va oferi ceva ce oamenii bogaţi au în ziua de azi: vieţi pline de tihnă şi distracţie.

Poate se va numi A Doua Renaştere, pentru că roboţii vor fi cei care vor trudi, în timp ce noi ne vom concentra pe chestiuni mai importante decât simplele sarcini de a ne hrăni şi de a ne asigura că avem un acoperiş deasupra capului. Vom crea şi ne vom bucura de artă splendidă. Ne vom îngriji şi ne vom dezvolta comunităţile şi ne vom implica în dezbateri politice viguroase şi sănătoase. Şi vom proteja şi ne vom bucura de planeta noastră frumoasă. Unii dintre noi s-ar putea folosi de tot acest timp ca să devină chiar oameni de ştiinţă amatori. Vom descoperi mai multe minuni ale universului – fără îndoială, cu ajutorul asistenţilor roboţi. La fel ca la precedenta Renaştere, aceasta va fi o perioadă în care cunoştinţele cu privire la locul nostru

în univers se vor extinde rapid.

Una peste alta, vom duce o viaţă bună. Preţul bunurilor de bază va scădea dramatic, pentru că vor fi produse de maşinării mai eficiente. Sărăcia va fi o amintire din trecutul îndepărtat – aşa cum este acum pentru cei din lumea dezvoltată, care îşi amintesc vag de adăposturile pentru săraci din epoca victoriană. Iar cei care vor alege să mai lucreze încă vor avea weekenduri de patru sau cinci zile. Cu toţii vom împărtăşi bogăţia creată de automatele care nu vor dormi niciodată.

Privind în urmă din anul 2062, anul 2000 va părea de modă veche. Trebuie să ne aducem aminte şi de cât de dramatic s-au schimbat ţările dezvoltate cum ar fi Australia, Statele Unite ale Americii şi Marea Britanie între anii 1900 şi 1962. Este posibil ca vieţile noastre să se schimbe la fel de mult între anii 2000 şi 2062.

În anul 1900, străzile erau pline de cai şi de căruţe. Mulţi oameni munceau în condiţii jalnice, făcând treburi repetitive şi obositoare. Speranţa de viaţă pentru cineva din acea vreme era de doar 41 de ani. Peste 62 de ani eram deja în drum spre Lună, străzile erau pline de maşini şi de camioane, iar condiţiile de muncă se îmbunătăţiseră extrem de mult. Epoca avioanelor a început, iar lumea s-a micşorat din ce în ce mai mult. Era o perioadă optimistă, după oroarea celor

două războaie mondiale şi după Marea Depresie. Ştiinţa aducea beneficii vizibile în vieţile noastre şi exista sentimentul că societatea avea să evolueze nespus de mult. Perioada cunoscută sub denumirea de *swinging sixties*[43] era pe aproape. Speranţa de viaţă se dublase, ajunsese la 71 de ani.

Anul 2062 ar putea fi o perioadă la fel de optimistă. Poate că vom fi suferit 50 de ani acomodându-ne cu noile tehnologii, dar roboţii vor prelua aproape toată munca şi viaţa se va îmbunătăţi incredibil de mult. Sărăcia va fi eliminată. Speranţa de viaţă va ajunge la o sută de ani, sau chiar mai mult. Anul 2000 va părea într-adevăr o epocă îndepărtată.

Jumătate din locurile de muncă

Una dintre temerile pe care le au economiştii este că multe locuri de muncă vor dispărea pur şi simplu până în anul 2062. O mare parte din munca noastră va fi făcută de roboţi. Acesta să fie sfârşitul muncii? Dacă maşinile ne vor depăşi, atât mental, cât şi fizic, noi, oamenii, ce vom mai face?

Aceasta nu este o temere minoră. În octombrie 2017, Google a anunţat că va cheltui un miliard de dolari în

[43] Revoluţie culturală a tinerilor în a doua parte a deceniului al şaptelea din secolul XX, în Anglia, care a pus accentul pe modernitate şi pe distracţie, (n. tr.).

următorii cinci ani prin programul „Grow with Google”, menit să-i ajute pe angajaţii din America să se reconvertească profesional, să sprijine afacerile mici şi locale să crească şi să susţină organizaţiile nonprofit din toată lumea active în aceste domenii. Angajaţii Google vor face muncă de voluntariat de un milion de ore pentru aceste organizaţii nonprofit.

Temerile cu privire la pierderea locurilor de muncă datează încă din anul 2013, când a fost făcut un studiu al impactului automatizării de către Carl B. Frey şi Michael A. Osborne, la Universitatea din Oxford. Studiul a făcut o predicţie care a ajuns să fie discutată foarte mult, cum că 47% din locurile de muncă din Statele Unite aveau să fie ameninţate de automatizare în următoarele două decenii. Au fost şi alte studii mai recente şi mai detaliate care au făcut predicţii la fel de dramatice.

Trebuie menţionat faptul că astfel de predicţii au fost făcute de economişti la intervale regulate de-a lungul ultimului secol. În anul 1930, spre exemplu, John Maynard Keynes a avertizat că va apărea „o nouă boală despre care mulţi cititori probabil că nu au auzit, dar despre care vor auzi cu prisosinţă în anii ce vor urma: şi anume şomajul tehnic”. Economiştii tind să se înşele destul de des, aşa că probabil nu ar trebui să ne surprindă faptul că aceste predicţii anterioare nu au

devenit realitate. În ciuda avertismentului lui Keynes, în ziua de azi şomajul are cele mai mici cote din istorie, în majoritatea ţărilor. Şi asta în pofida faptului că populaţia lumii nu a fost niciodată atât de numeroasă în toată istoria ei. Munca nu s-a terminat – într-adevăr, se pare că a mai rămas mult de lucru. Şi mulţi dintre noi ajungem să petrecem din ce în ce mai mult timp muncind.

Sunt două lucruri care ies în evidenţă din raportul lui Frey şi Osborne. În primul rând, raportul în sine a fost automatizat în mare parte. Autorii au folosit învăţarea automată pentru a prezice cu precizie ce slujbe pot fi eliminate din 700 de profesii diferite. Bineînţeles, este ironic că un raport despre automatizarea muncii a fost chiar el automatizat; se pare că a trebuit să aşteptăm ca maşinile să ne spună că o să fim automatizaţi, înainte să fim pregătiţi să luăm în serios această posibilitate.

Al doilea lucru care iese în evidenţă din raportul lui Frey şi Osborne este faptul că procentul de locuri de muncă considerate a fi în pericol din cauza automatizării este nemaipomenit de mare. Conform predicţiei, aproape jumătate din slujbe vor fi expuse riscului de a dispărea. Asta a dus la crearea unei meme care a putut fi repetată din nou şi din nou.

Andy Haldane, directorul economist de la Banca Angliei, a prevăzut în luna noiembrie a anului 2015 că

în jur de jumătate din slujbele din Marea Britanie vor fi în pericol de a fi automatizate. Iar în luna octombrie a anului 2016, Jim Yong Kim, preşedintele Băncii Mondiale, a prevăzut că 69% din locurile de muncă din India şi 77% din cele din China vor fi expuse riscului. Cu procente atât de mari, nu-i de mirare că mulţi dintre noi am început să ne îngrijorăm.

Înainte să mergem mai departe, să stabilim odată pentru totdeauna cum e cu procentul de locuri de muncă în pericol. Oricine susţine că poate prezice numărul de slujbe aflate în pericol în următoarele două decenii, mai ales cu o astfel de precizie, estimând un procent fix de 45%, 69% sau 77%, se păcăleşte singur, şi încearcă să te păcălească şi pe tine. Nu avem idee care este numărul real; sunt prea multe necunoscute.

Date de învăţare

Un motiv pentru care predicţii precum cea a lui Frey şi Osborne şi cea a Băncii Mondiale sunt cel mai probabil eronate este faptul că sunt erori în imput – în datele pe baza cărora a avut loc învăţarea. În informatică se foloseşte o expresie bine-cunoscută: bagi gunoi, iese gunoi.

În acest caz, predicţiile clasificatorului folosit pentru studiul lui Frey şi Osborne depind decisiv de datele pe

baza cărora a fost făcută învăţarea. Cercetătorii au clasificat manual dacă 70 din 702 slujbe diferite vor fi în pericol de a fi automatizate în următoarele două decenii. Clasificarea lor a fost binară: o slujbă va risca să fie automatizată sau nu. Dar unele dintre slujbe se încadrau mai degrabă undeva la mijloc. Spre exemplu, o ocupaţie pe care ei au clasificat-o ca fiind în pericol a fost aceea de *contabil şi revizor contabil.* Vor exista cu siguranţă anumite aspecte din profesia unui contabil şi revizor contabil ce vor fi automatizate în următoarele decenii, dar mă îndoiesc că slujba de contabil şi revizor contabil va dispărea.

În orice caz, Frey şi Osborne au clasificat 37 din cele 70 de slujbe la care s-au uitat chiar ei ca fiind în pericol. Adică, peste jumătate din datele de învăţare care au fost oferite ca imput clasificatorului erau slujbe despre care se credea că sunt la risc să fie automatizate. Prin urmare, nu este surprinzător că programul a prezis că aproximativ jumătate din setul complet de 702 de locuri de muncă erau expuse riscului de automatizare. Dacă setul lor de învăţare ar fi fost mai prudent realizat – să spunem că ar fi declarat că una din patru slujbe se află în pericol –, atunci este posibil ca predicţia per total cu privire la setul întreg de locuri de muncă să fi fost la fel de precaută.

În ianuarie 2017, am decis să explorez această idee

printr-un sondaj propriu, pe care l-am realizat în acelaşi timp cu cel legat de cât timp va mai dura până când maşinile vor ajunge la fel de inteligente ca oamenii. Am rugat 300 de experţi în IA şi robotică să clasifice care dintre slujbele din setul de învăţare al lui Frey şi Osborne se află în pericol de a fi automatizate în următoarele două decenii. Am rugat să facă acelaşi lucru şi 500 de non-experţi, cititori ai unui articol pe care l-am scris despre evoluţia roboţilor de poker. Non-experţii au emis aproape aceleaşi predicţii ca ale lui Frey şi Osborne. Dar experţii în IA şi robotică au fost mult mai prudenţi[44]. Ei au prezis cu 20% mai puţine locuri de muncă automatizate decât Frey şi Osborne. Bineînţeles, până şi acest procentaj ar însemna un număr semnificativ de profesii, dar nu este la fel de dramatic ca studiul de la Oxford.

Repararea bicicletelor

Chiar dacă presupunem că am putea corecta

[44] Economiştii au fost printre cei care au susţinut cu cea mai multă convingere că va exista pericolul şomajului tehnic. De aceea este ironic faptul că cei mai mulţi experţi şi non-experţi din sondajul meu nu au fost de acord cu posibilitatea ca slujba de economist să fie automatizată în următoarele două decenii. Patru din zece non-experţi au prezis că slujba de economist ar putea fi automatizată în următoarele două decenii, în timp ce unul din cei opt experţi a făcut această predicţie.

predicţiile lui Frey şi Osborne cu date de intrare mai bune, tot nu am rezolva problema. Spre exemplu, clasificatorul cu învăţare automată oferă o probabilitate de 94% pentru ca meseria de reparator de biciclete să fie automatizată în următoarele două decenii. Dar vă pot asigura că şansele sunt aproape nule ca această slujbă să fie automatizată până atunci, nici măcar cele mai mici părţi ale sale.

Această eroare scoate la iveală unele dintre limitele studiului lor. În primul rând, este dificil să repari biciclete, sunt greu de mânuit şi au multe componente care nu sunt standard. În al doilea rând, nu va fi viabil din punct de vedere economic să înlocuim un reparator de biciclete cu un robot. Nu este o slujbă plătită prea bine şi va necesita un robot foarte scump. Şi în al treilea rând, repararea de biciclete este o profesie socială. Include conversaţia cu clientul, vânzarea ultimei truse apărute, sfaturi cu privire la locurile potrivite pentru mersul pe bicicletă. Nu este vorba doar despre reparatul bicicletelor.

Clasificatorul folosit de Frey şi Osborne nu a luat în considerare dacă va fi fezabil din punct de vedere economic să se automatizeze o slujbă. A ţinut cont doar de faptul că automatizarea poate fi posibilă din punct de vedere tehnic. În mod similar, clasificatorul lor nu s-a gândit la faptul că am putea *prefera* ca un om să aibă

acea profesie. Şi mai sunt o serie de alte aspecte de care datele lor nu au ţinut cont şi care vor decide ce slujbe vor fi automatizate.

Avioanele

O altă predicţie din studiul lui Frey şi Osborne a fost că exista o probabilitate de 55% ca ocupaţia de *pilot comercial* să fie automatizată în următoarele două decenii. Aceasta este o predicţie interesantă deoarece, dintr-o perspectivă tehnică, deja putem automatiza în mare parte această ocupaţie în prezent. De fapt, de cele mai multe ori un calculator conduce avionul. (Când aterizarea nu e atât de lină, ştii că pilotul uman a fost cel care controla avionul.) Mai mult de atât, spaţiul aerian este un mediu foarte controlat, ceea ce face mai uşoară automatizarea pilotării avioanelor, în comparaţie cu cea a automobilelor.

Nu este clar ce înseamnă această probabilitate de 55%. Cu siguranţă, e puţin probabil ca, în 20 de ani, 55% din piloţii comerciali să fie înlocuiţi de calculatoare. De fapt, Boeing prevede că peste două decenii vor fi necesari 600.000 de piloţi noi. Ar putea fi în interesul lor să spună asta, dar oricum, această ocupaţie nu pare mai deloc în pericol. Judecând după clasa mijlocie în creştere din China şi din India, care

începe să îşi dezvolte apetitul pentru a călători, cererea pentru piloţi este în creştere. Şi automatizarea pare să aibă un impact foarte mic asupra acestei cereri.

Mă aştept să fie nevoie de mai mult de două decenii pentru ca piloţii comerciali să fie înlocuiţi de calculatoare. Atunci când călătorim ne place ideea că avem o persoană în faţă care conduce, a cărei viaţă va avea de suferit dacă lucrurile nu merg cum trebuie, împreună cu ale noastre. Cel mai probabil vor trece mai bine de 20 de ani până ce va deveni acceptabil din punct de vedere social ca pilotul uman să fie lăsat în urmă. Chiar dacă majoritatea accidentelor din ziua de azi sunt provocate de greşelile piloţilor.

Cu câteva decenii în urmă, când m-am urcat la bordul unui avion, mi-am dorit ca acea persoană din faţă să fie un pilot cu păr alb şi multă experienţă în spate. Se spune că există piloţi bătrâni şi piloţi curajoşi, dar niciun pilot bătrân şi curajos. Dar azi îmi doresc să văd un tocilar tânăr care ştie cum să controleze toate ecranele. Aşa că slujba va suferi modificări. Va ţine mai mult de folosirea computerelor decât de zborul de modă veche. Dar cel mai probabil, cererea pentru piloţi umani nu va fi în scădere.

Fotomodele roboţi

O altă predicţie din studiul lui Frey şi Osborne era că exista o probabilitate de 98% ca fotomodelele din lumea modei să fie automatizate în următoarele două decenii. Chiar aşa? O să înlocuim manechinele umane, cu buzele lor ţuguiate, cu felul în care îşi mişcă şoldurile şi cu mersul ţanţoş, cu nişte roboţi care să meargă pe podiumurile lumii?

Din nou, mă îndoiesc că aşa vor sta lucrurile. Nu vrem să ştim cum arată hainele pe un robot, ci pe un om. Nu aspirăm să fim un robot care poartă un ceas scump, ci să fim acel bărbat model care arată ca un om avut. Şi să nu uit de o obiecţie tehnică: roboţii nu vor reuşi prea curând să meargă pe tocuri.

Există un loc în care modelele umane vor fi înlocuite, şi acela este lumea virtuală. Vom fi capabili să creăm modele în mod digital. Acestea nu se vor putea distinge de versiunea lor umană. Şi îşi vor face treaba pentru mult mai puţin de 10.000 de dolari pe zi.

Industria muzicală sugerează în ce direcţie s-ar putea îndrepta ocupaţia de manechin. Muzica digitală nu a avut ca rezultat mai puţini muzicieni umani. Din contră, Departamentul Muncii al SUA a prezis că cererea pentru muzicieni nu va scădea, ci va creşte puţin în următorul deceniu. Industria muzicii s-a mutat în lumea analogă. Mulţi cântăreţi fac mai mulţi bani acum din concerte decât din înregistrări. Valorizăm

experienţele – vrem să-i auzim pe idolii noştri muzicali în persoană. Într-un fel asemănător, ne putem aştepta ca modellingul să rămână în continuare o ocupaţie pentru oameni, dar să se axeze mai mult pe prestaţia în cadrul unor evenimente. Şi până şi în fotografia de modă va exista o trecere de la artificial la real. Departamentul Muncii al SUA a prezis că numărul de persoane angajate ca modele nici nu va creşte, nici nu va scădea semnificativ în următorul deceniu.

Numărul real de locuri de muncă aflate în pericol

Clasificatorul lui Frey şi Osborne nu a luat în considerare şi multe variabile tehnice, economice şi sociale. În mod nesurprinzător, clasificatorul lor a prezis că slujbele plătite mai puţin riscă să fie automatizate, dar nu este viabil din punct de vedere economic să fie automatizate. Chiar dacă presupunem că am extinde modelul pentru a ţine cont de aceşti factori importanţi şi de mulţi alţii, predicţia lor cum că 47% din slujbe riscă să fie automatizate nu este echivalentă cu faptul că 47% dintre oameni vor fi şomeri. Vor exista multe slujbe noi create de tehnologie, şi mai sunt şi alţi factori care trebuie luaţi în considerare, cum ar fi schimbările demografice care au loc.

Istoria tehnologiei ne-a arătat că noile tehnologii au creat mai multe locuri de muncă faţă de câte au distrus. Înainte de revoluţia industrială, cei mai mulţi oameni lucrau în agricultură sau erau meşteşugari. Multe dintre aceste slujbe au fost mecanizate. Dar s-au creat multe profesii noi, în birouri şi în fabrici. În mod similar, ne putem aştepta ca odată ce IA automatizează slujbele vechi, să fie create unele noi.

Mesajul este că orice predicţie cu privire la numărul de locuri de muncă aflate în pericol din cauza automatizării este inerent nesigură, la fel ca impactul asupra nivelului de angajare. Nu pare probabil ca 47% din slujbe să ajungă să fie automatizate. Adevărul este că nimeni nu poate spune care este efectul net asupra angajărilor.

Ceea ce ştim

Există dovezi că *unele* ocupaţii încep să fie automatizate – şi că unele dintre aceste ocupaţii nu vor fi înlocuite de altele proaspăt create. Un studiu MIT din anul 2017 a analizat impactul automatizării în Statele Unite între anii 1993 şi 2007. A ajuns la concluzia că roboţii industriali au redus numărul de locuri de muncă per total.

În medie, fiecare robot nou a înlocuit aproximativ 5, 6

muncitori. Şi nu s-a observat o creştere compensatorie în alte ocupaţii. De fapt, studiul a estimat că fiecare robot în plus, la 1000 de muncitori, reducea populaţia totală de persoane angajate din SUA cu un procent de 0,34%.

Automatizarea a pus presiune şi pe slujbele rămase. Fiecare robot suplimentar la 1000 de angajaţi reducea salariile cu 0,5%. Pe parcursul celor 14 ani, numărul de roboţi industriali din Statele Unite s-a cvadruplat, eliminând ceea ce s-a estimat a fi în jur de 500.000 de locuri de muncă.

Industria petrolieră prezintă un studiu de caz instructiv cu privire la dimensiunile problemei. Preţul petrolului a scăzut de la 115 dolari pe baril în august 2014 la sub 30 de dolari la începutul lui 2016. Acest lucru a determinat industria să reducă numărul de angajaţi şi să introducă mai multă muncă automatizată. Aproape jumătate de milion de locuri de muncă au dispărut din industria petrolieră la nivel global. Dar acum, în timp ce preţul petrolului îşi revine, şi industria creşte din nou, mai puţin de jumătate dintre aceste locuri de muncă au revenit. Automatizarea a redus numărul de 20 de oameni care erau angajaţi în medie pentru o sondă de petrol la doar cinci.

Slujbe închise versus slujbe deschise

Un motiv pentru care automatizarea nu va elimina anumite locuri de muncă este că, în unele cazuri, ne va permite să facem mai multe la acele locuri de muncă. În cazul acesta, este important să facem distincţia dintre slujbe „deschise” şi „închise”. Automatizarea va tinde să sporească slujbele deschise, dar să le înlocuiască pe cele închise.

Slujbele închise sunt acelea în care există o cantitate fixă de lucru. Spre exemplu, curăţarea geamurilor este o profesie închisă. Există doar un număr limitat de geamuri care pot fi curăţate pe planetă. La expoziţia de calculatoare CeBIT din anul 2017 am văzut un număr de prototipuri de roboţi curăţători de geamuri. Odată ce roboţii vor putea curăţa geamurile, lucru care se va petrece destul de curând, slujba unui curăţător de geamuri uman va dispărea. Cel puţin va dispărea din ţările dezvoltate, unde este scump să plăteşti curăţătorii de geamuri, iar aceştia sunt în pericolul de a cădea de la înălţime.

La fel şi repararea de biciclete este o slujbă închisă. Am exprimat deja dubii cu privire la posibilitatea automatizării acestei slujbe, dar chiar dacă va fi automatizată, nu e ca şi cum vor fi dintr-odată mai multe biciclete de reparat.

Prin comparaţie, slujbele deschise se vor extinde pe măsură ce vor deveni automate. Spre exemplu, jobul de chimist reprezintă o meserie deschisă. Dacă eşti chimist şi ai nişte instrumente care îţi automatizează treaba, asta te ajută doar să faci şi mai multă chimie. Poţi depăşi barierele actuale cu privire la înţelegerea chimiei mult mai rapid. Cel mai probabil nu vei rămâne fără lucruri noi de înţeles în chimie.

Desigur, majoritatea profesiilor nu sunt nici complet deschise, nici complet închise. Să luăm ca exemplu profesiile juridice. Din moment ce calculatoarele preiau din ce în ce mai mult sarcinile legale de rutină, costul accesului la servicii legale va scădea. Acest lucru va extinde piaţa pentru avocaţi, generând o cerere şi mai mare, şi astfel, cu toţii vom avea acces la consiliere juridică şi mai bună. Cel mai probabil va fi şi mai mult de lucru pentru avocaţii cu experienţă. Dar este greu să ne închipuim că vor exista multe slujbe juridice de nivel intermediar. S-ar putea să devină mai greu pentru tinerii absolvenţi să intre în competiţie cu avocaţii roboţi, care au citit deja întreaga literatură ce ţine de drept, care nu au niciodată nevoie de somn, nu greşesc şi nu au nevoie de un salariu.

Meserii automatizate parţial

Un argument pentru a susţine că automatizarea a 47% din slujbe nu va duce la şomaj în masă este că doar *anumite părţi* din aceste 47% din slujbe vor fi automatizate. Nu sunt de acord cu acest argument. Dacă automatizezi o slujbă, atunci vei putea să faci aceeaşi treabă cu mai puţini oameni.

Să ne gândim din nou la profesia de avocat. La o analiză mai atentă asupra timpului pe care-l petrec avocaţii pentru diverse aspecte ale muncii lor, am observat că doar în jur de un sfert din timp îl petrec îndeplinind sarcini ce ar putea fi automatizate în viitorul apropiat [45] . Haide să lăsăm deoparte deocamdată faptul că multe dintre sarcinile rămase ar putea să fie automatizate în viitorul îndepărtat. În afara cazului în care creăm mai multe îndatoriri legale, atunci am putea să ne ocupăm de sarcinile juridice cu o treime din avocaţii pe care-i avem în ziua de azi. Avocaţii ar putea să evolueze şi să folosească timpul în plus ca să lucreze mult mai bine. Dar unele firme juridice pur şi simplu vor scădea preţurile cu trei sferturi şi vor da afară un sfert din angajaţii lor ca să compenseze pentru salariul lor redus.

Acest argument – doar anumite părţi dintr-o meserie vor fi automatizate – a fost folosit ca să susţină şi că una

45

dintre profesiile care par să se afle în pericolul cel mai mare de a fi automatizate este de fapt în siguranţă. Şoferii de tir nu trebuie să-şi facă griji, s-a susţinut, deoarece vor fi întotdeauna cazuri în care maşinile nu ar putea face faţă singure. Tirul ajunge într-un şantier unde un angajat îi face un semn din mână. Tirul trebuie să conducă în jurul unei fabrici care nu se află pe nicio hartă GPS. Tirurile autonome pur şi simplu nu vor putea să se descurce în asemenea situaţii.

Vestea proastă pentru şoferii de tir este că argumentul face abstracţie de condusul de la distanţă. Companii cum ar fi Starsky Robotics deja testează tiruri autonome asupra cărora şoferii preiau controlul de la depărtare atunci când maşina nu mai poate face faţă. Un astfel de şofer de la distanţă va putea să aibă grijă de mai multe tiruri autonome. Aşa că, deşi vom mai avea şoferi umani pentru o perioadă, vor fi mult mai puţini pe viitor decât acum.

A munci mai puţin

Un alt motiv pentru care automatizarea de 47% nu va duce la un şomaj de 47% este că s-ar putea să avem o săptămână de lucru mai scurtă. Aşa s-a întâmplat în timpul revoluţiei industriale. Înainte de ea, mulţi oameni se trezeau cu noaptea-n cap, se duceau pe

terenuri şi munceau până apunea soarele, după care se culcau. Mulţi munceau în jur de 60 de ore pe săptămână. După revoluţia industrială, munca s-a redus la 48 de ore pe săptămână pentru majoritatea. Şi unii dintre noi chiar beneficiem de săptămâni întregi de concediu în fiecare an.

La fel s-ar putea întâmpla şi cu revoluţia IA. Am putea scurta ziua de lucru. Sau am putea avea un weekend de trei sau de patru zile. Dar acest lucru ar presupune ca o parte din banii obţinuţi în urma creşterii productivităţii să fie împărţiţi cu toată lumea, şi deocamdată nu prea pare să se întâmple acest lucru. Ba chiar pare să se întâmple opusul. Creşterea salariilor pentru majoritatea angajaţilor din economiile dezvoltate s-a oprit, sau nu mai poate face faţă inflaţiei limitate cu care avem de-a face.

Există un argument care spune că ne putem permite să muncim mai puţin. IA va scădea preţul multor necesităţi de bază. Maşinării mai eficiente vor face ca multe dintre lucrurile esenţiale în viaţă să fie mai ieftine. Aşa că s-ar putea să nu fie nevoie de creşterea salariilor pentru a lucra mai puţin. Am putea – măcar în teorie – să trăim cu mai puţin.

Răsplătirea tuturor formelor de muncă

Pe lângă faptul că vom munci mai puţin, am putea să şi plătim oamenii mai mult pentru treaba pe care o fac. Mulţi dintre noi au grijă de rudele mai în vârstă, îşi cresc copiii şi fac muncă de voluntariat în comunităţi. Aceasta este o muncă vitală pentru societate, dar adesea nu este plătită. Pe măsură ce societatea va deveni mai bogată, vom putea găsi căi prin care să îi răsplătim pe oameni pentru astfel de lucruri. Să le apreciem mai mult. Mulţi dintre noi ne obţinem statutul în societate în funcţie de munca plătită pe care o facem. Ne considerăm persoane harnice, care-şi plătesc taxele, membri productivi ai comunităţilor noastre. Trebuie să îi respectăm şi pe cei care fac acea muncă neplătită, care este vitală pentru ca societatea să continue să funcţioneze aşa cum trebuie.

Iar asta scoate în evidenţă o problemă mai mare. Sunt multe slujbe plătite şi neplătite pe care ar trebui să le apreciem mai mult, printre care cele ale profesorilor, asistentelor, poliţiştilor şi poliţistelor, pompierilor şi mulţi alţii fără de care nu am putea trăi. Bonusul de productivitate pe care îl introduce IA ar putea fi investit, dacă alegem să facem asta, în moduri care recunosc meritele acestor roluri şi le răsplătesc.

O planetă care îmbătrâneşte

De asemenea, trebuie să ţinem cont şi de datele demografice în permanentă schimbare. Spre exemplu, în multe ţări dezvoltate oamenii au mai puţini copii şi trăiesc mai mult. Drept consecinţă, vârsta de pensionare este în creştere, dar în aceeaşi măsură în care creşte şi speranţa de viaţă.

Colegul meu Rodney Brooks, unul dintre cei mai faimoşi experţi contemporani în robotică, a spus că el nu îşi face griji în privinţa şomajului tehnic. Din contră. El este de părere că roboţii vor ajunge exact la timp ca să ne salveze atunci când ne vom afla într-o cumpănă. Fără roboţi, nu va mai rămâne nimeni care să facă toată treaba ce îi va susţine pe toţi oamenii retraşi din activitate.

Japonia este unul dintre locurile în care această schimbare este foarte importantă. Japonezii au cea mai lungă speranţă de viaţă de pe planetă, iar companiile japoneze investesc foarte mult în construirea roboţilor îngrijitori. Societatea niponă pare să primească cu braţele deschise ideea folosirii roboţilor care să îi îngrijească pe cei în vârstă, iar restul lumii s-ar putea să le urmeze exemplul în cele din urmă. Dacă Rodney Brooks are dreptate, s-ar putea să nu avem de ales. Dar să presupunem că avem de ales, nu sunt acestea nişte meserii pe care am dori să le păstrăm pentru oameni? Chiar vrem să fim îngrijiţi de roboţi la bătrâneţe?

Învingători şi învinşi

Tehnologiile tind să afecteze grupuri diferite în moduri diferite, iar IA va face cu siguranţă acelaşi lucru. Unele grupuri vor fi avantajate, iar altele dezavantajate. Atunci, cine vor fi învingătorii, şi cine învinşii?

Un grup care pare deja să piardă este format din bărbaţii în jur de 20 de ani fără studii superioare. În anul 2015, 22% din bărbaţii americani cu vârste cuprinse între 21 şi 30 de ani, şi fără diplomă de licenţă, nu au lucrat deloc în ultimele 12 luni. Acest grup demografic era înainte baza forţei de muncă. Era cel mai muncitor grup de angajaţi din America, pe care te puteai baza cel mai mult. Se alăturau clasei muncitoare imediat ce terminau şcoala şi lucrau încontinuu până la pensie. Azi, unul din cinci nu are serviciu şi nici nu îşi caută de lucru.

Fără slujbe, acest grup este mai puţin probabil să se căsătorească, să plece de acasă sau să se implice în politică. Mulţi stau acasă, pierduţi în lumi virtuale. Rata mortalităţii a crescut în cazul lor deoarece mulţi îşi caută refugiul în băutură şi în droguri. Dacă nu pot să urce în ierarhia angajaţilor, atunci nu vor avea niciodată o slujbă decentă? Viitorul pare destul de

sumbru în cazul lor.

O mare de tipi

Un alt grup care este în pierdere îl constituie femeile. Deja avem un termen pentru asta: „marea de tipi”. Termenul a fost introdus în anul 2016 de Margaret Mitchell, pe atunci cercetătoare IA la Microsoft Research, iar acum la Google, şi subliniază faptul că doar 10% din cercetătorii din domeniul IA sunt femei. Acest dezechilibru al sexelor dăunează progresului dezvoltării IA. Unele probleme fundamentale nu vor fi abordate pentru că nu se pun niciodată întrebările de bază. Nu ar trebui o aplicaţie medicală să ţină cont de ciclul menstrual al unei femei?[46] Atunci când se fac angajări, cum putem elimina prejudecăţile împotriva celor care îşi iau concediu parental?

În general, femeile vor ieşi în pierdere pentru că slujbele nou create vor implica tehnologia. Dacă dezechilibrul dintre sexe din cadrul meseriilor ce ţin de tehnologie va persista, atunci femeile vor avea un dezavantaj în ceea ce priveşte angajarea în joburi noi. Pe de altă parte, vor fi mai mulţi bărbaţi care ocupă posturi ce sunt în pericol de a fi automatizate, spre

[46] Acest exemplu a fost ales cu atenţie. Lansarea iniţială a HealthKit pentru iWatch al Apple din iunie 2014 nu urmărea ciclurile menstruale.

deosebire de femei. Nu este clar care dintre aceşti doi factori aflaţi în concurenţă va fi cel decisiv.

În prezent, femeile nu sunt singurul grup mare de persoane care au de suferit din punctul de vedere al locurilor de muncă. Alte grupuri, cum ar fi persoanele de culoare şi cele hispanice, sunt de asemenea slab reprezentate în tehnologie în general, şi în cercetarea în IA în special. Din nou, este posibil ca acest lucru să aibă un impact negativ asupra drepturilor care au fost câştigate foarte greu, cum ar fi egalitatea rasială. Dacă printre cei care vor construi IA lipseşte diversitatea, va fi greu să ne asigurăm că aceste sisteme vor fi imparţiale.

Nu există nicio soluţie simplă. Ştim, spre exemplu, că dezechilibrul dintre sexe apare la o vârstă mică, atunci când fetele încep să aleagă materii la şcoală care nu includ matematica, ştiinţa sau tehnologia. Dar recunoaşterea acestor probleme este măcar un prim pas pe drumul spre un viitor mai imparţial.

Lumea în curs de dezvoltare

Un ultim grup foarte mare care s-ar putea să aibă de suferit este reprezentat de cei din ţările care se află în curs de dezvoltare. În lumea dezvoltată, am beneficiat de revoluţia industrială şi ne-am îmbunătăţit calitatea

vieţii. Dar nu este deloc clar că beneficiile revoluţiei IA vor ajunge şi în lumea în curs de dezvoltare.

Majoritatea profitului obţinut în urma transformărilor tehnologice merge la companiile mari de tehnologie. Bogăţia lor nu este împărţită prea bine nici cu restul lumii dezvoltate, darămite cu lumea în curs de dezvoltare. În timp ce în multe locuri sărăcia este în scădere, inegalitatea atât în interiorul ţărilor, cât şi între ele este în creştere.

Globalizarea a externalizat multe meserii de la lumea dezvoltată la cea în curs de dezvoltare. Centrele telefonice s-au mutat în India şi în alte locuri. Producţia s-a mutat dintr-o ţară în alta, în locuri unde muncitorii erau plătiţi cu mai puţin. Lanţurile de aprovizionare au devenit globale. Acestea au fost nişte veşti bune pentru ţările în curs de dezvoltare.

Acum, s-ar putea ca aceste tendinţe să înceapă să se inverseze. Eficienţa crescută pe care o vor aduce IA şi robotica va face ca multe dintre slujbe să se mute iar la fabrici automatizate şi la birouri în lumea dezvoltată. Lanţurile de aprovizionare vor fi mai scurte (şi astfel mai rezistente la perturbările produse de climă). Oare toate acestea vor împiedica ţările în curs de dezvoltare să înflorească? Unii speră că lumea în curs de dezvoltare va putea să sară peste construcţia de infrastructură scumpă şi instituţii, lucruri care au

acompaniat industrializarea în lumea dezvoltată. Nu este nevoie să creezi un serviciu poştal când poţi foarte bine să trimiţi un e-mail; nu este nevoie să pui cabluri în toată ţara când poţi să treci direct la 5G.

Şi totuşi, s-ar putea să nu fie de ajuns pentru ca lumea în curs de dezvoltare să prospere. Este o luptă continuă pentru forţarea marilor companii farmaceutice să ofere medicamente într-un mod accesibil celor aflaţi în ţările în curs de dezvoltare. Lucrurile nu stau prea bine pentru aceştia din urmă, care speră că vor putea obţine acces la proprietatea intelectuală care le va permite şi lor să fie un producător activ de IA, şi nu doar un consumator.

Şoferii de camioane şi de taxi

În lumea dezvoltată, cei care se află în cel mai mare pericol sunt probabil şoferii. Cei care conduc taxiuri, camioane şi maşini de livrări vor fi poate primele victime colaterale ale revoluţiei IA. Peste trei milioane de oameni din Statele Unite ale Americii sunt angajaţi ca şoferi. Multe dintre locurile lor de muncă sunt expuse riscului de a dispărea în următoarele două decenii. Tesla avea ca obiectiv obţinerea autonomiei totale („nivelul 5”) pentru maşinile lor, până în anul 2019 – dar trebuie menţionat faptul că Tesla nu s-a

descurcat prea bine până acum să îşi respecte termenele-limită. Atât Volvo, cât şi Ford au anunţat că până în anul 2021 vor vinde şi ei maşini cu pilot automat complet. Maşinile fără şofer par să fie la o aruncătură de băţ.

Siguranţa va conduce toate acestea, ca să fac o glumă. Invenţia automobilului ne-a oferit mobilitate personală, dar am avut un preţ foarte mare de plătit. Peste 37.000 de oameni au murit pe drumurile din Statele Unite în anul 2016. Peste 1000 au murit pe drumurile din Australia în acel an. Şi în jur de 95% au fost accidente provocate de greşeala şoferului. Adevărul este că suntem nişte şoferi groaznici. Conducem când suntem beţi. Conducem când suntem obosiţi. Suntem distraşi de telefonul mobil. Încălcăm regulile rutiere. Depăşim atunci când nu trebuie. Trecem pe roşu. În Statele Unite există o probabilitate de unu la o sută să mori într-un accident ce implică un vehicul cu motor[47]. Accidentele rutiere sunt prima cauză a morţii în rândul adolescenţilor americani, fiind chiar mai periculoase decât armele de foc. Cu cât vom înlocui mai repede şoferii umani cu calculatoare pe care ne putem baza, cu

[47] În Statele Unite, există o probabilitate de 1 la 114 pe parcursul vieţii să mori într-un accident rutier. Ca să punem asta în perspectivă, există o probabilitate de 1 la 370 să fii ucis de o armă de foc şi o probabilitate de unu la 9800 să mori într-un accident de avion. Poate că ar trebui să fii mai îngrijorat din cauza maşinilor, nu a avioanelor.

atât mai bine. În anul 2062 ne vom uita în urmă cu stupoare, gândindu-ne la masacrul care am permis să aibă loc cândva pe străzile noastre.

Un alt factor decisiv va fi costul. În jur de trei sferturi din costurile transportului de bunuri cu camionul este reprezentat de salariile şoferilor. Cea mai scumpă parte din călătoria ta cu Uberul o reprezintă şoferul. Uber deja testează maşini care să nu aibă nevoie de şofer. Este singura cale prin care compania poate să se extindă precum alte companii tehnologice, cum ar fi Google şi Facebook, justificând astfel sumele enorme la care sunt evaluate.

Câştigurile obţinute în urma maşinilor autonome vor fi impresionante. Un camion autonom nu va avea nevoie să ia pauze ca să conducă pe o distanţă de două ori mai lungă, la aproape un sfert din costuri. Asta înseamnă o creştere a profitului de aproximativ opt ori. Şoferii umani pur şi simplu nu vor putea să concureze cu asta. Şofatul nu va mai fi o aptitudine pentru care oamenii vor fi plătiţi pe viitor.

În cazul şoferilor de camioane, tranziţia ar putea fi destul de uşor de suportat. Condusul camioanelor este o profesie dominată de oamenii în vârstă în Australia: vârsta medie a unui şofer de camion este de 47 de ani. Peste aproximativ un deceniu, mulţi vor ieşi la pensie, iar ocupaţia lor va fi preluată de camioane automate.

Tinerii nu vor alege această slujbă.

Pentru şoferii de taxi, tranziţia s-ar putea să se petreacă mai rapid şi mai dureros. Pentru majoritatea populaţiei va fi un lucru grozav, întrucât preţul unui Uber va scădea dramatic. Dar pentru şoferii de Uber, viitorul pare sumbru. Este o oarecare ironie în faptul că una dintre cele mai noi meserii de pe planetă – şoferul de Uber – ar putea fi şi una dintre meseriile cu cea mai scurtă durată de viaţă.

Va trebui să avem de-a face cu toate efectele secundare ale vehiculelor autonome. Ce se va întâmpla cu toate opririle pentru tiruri când nu vor mai exista şoferi care să aibă nevoie să facă o oprire, să doarmă şi să mănânce? Ce se va întâmpla atunci cu preţurile proprietăţilor imobiliare din suburbii şi din zonele rurale? Automobilul a fost cel care a definit America. În curând, vehiculele autonome vor fi cele care o vor redefini.

Învăţare pe parcursul întregii vieţi

O modalitate prin care vom putea fi cu un pas înaintea computerelor va fi învăţarea unor noi aptitudini pe măsură ce vor fi inventate tehnologii noi. Mulţi dintre noi vor trebui să se reinventeze în mod repetat. Noile tehnologii vor crea locuri de muncă noi.

Dar acestea vor pretinde alte aptitudini faţă de profesiile care au fost înlocuite. Învăţarea nu va fi ceva ce vom face doar pe timpul şcolii sau universităţii. Va trebui să învăţăm toată viaţa.

Vor fi necesare nişte schimbări foarte mari în sistemul educaţional. Cum îi vom ajuta pe oameni să înveţe aptitudini noi când ei deja vor face parte din forţa de muncă? Cum le vom oferi copiilor abilitatea de a învăţa abilităţi noi mai târziu în viaţă? Şi cum pot angajatorii şi guvernele să susţină învăţarea în timpul lucrului?

Inteligenţa artificială ar putea să fie atât problema, cât şi rezolvarea. IA creează problema deoarece îi lasă pe oameni fără serviciu, ceea ce îi obligă să se reprofileze. Dar IA îi poate şi ajuta să înveţe. De exemplu, poate construi instrumente care să susţină învăţarea personalizată.

Va trebui să luăm în considerare modificări majore în privinţa programei. Cele mai importante aptitudini din viitor nu sunt cele tehnologice. Anumite abilităţi tehnice vor fi lăsate în urmă. Cu siguranţă, mai multe „materii ŞTIM” (ştiinţă, tehnologie, inginerie şi matematică) nu reprezintă răspunsul. Va exista o cerere limitată pentru programatori de computer; când IA va atinge noi culmi, mare parte din programare va fi făcută chiar de calculatoare.

Oamenii vor avea nevoie, în schimb, de aptitudini analitice puternice. Vor avea nevoie de inteligenţă emoţională şi socială. Şi vor avea nevoie de toate celelalte calităţi care ne fac oameni – creativitatea, rezistenţa, determinarea şi curiozitatea. Aceste însuşiri umane ne vor ajuta să fim mai presus de calculatoare.

Meserii noi

Toate tehnologiile creează slujbe noi în aceeaşi măsură în care le şi distrug. Aşa s-a întâmplat în trecut, şi nu avem niciun motiv pentru care să presupunem că nu se va întâmpla şi pe viitor. Totuşi, nu există nicio lege fundamentală a economiei care spune că numărul de locuri de muncă distruse este egal cu numărul de locuri de muncă create. În trecut, multe slujbe au fost create, iar apoi distruse, dar nu trebuie să fie la fel şi acum. De data aceasta, ar putea fi diferit.

În timpul revoluţiei industriale, maşinăriile au preluat multe sarcini fizice de care se ocupau oamenii. Dar tot ne-au rămas sarcinile cognitive. De data aceasta, pe măsură ce maşinile încep să ia asupra lor şi multe dintre sarcinile cognitive, apare o întrebare îngrijorătoare: oamenilor ce le-a mai rămas?

Unul dintre colegii mei a sugerat că vor exista multe meserii noi, cum ar fi cea de reparator de roboţi. Nu

sunt deloc convins de asta. Miile de oameni care obişnuiau să vopsească şi să sudeze în fabricile de maşini au fost înlocuiţi de doar câţiva reparatori de roboţi. De asemenea, poate că roboţii vor ajunge să se repare unii pe alţii. Deja avem fabrici în care roboţii creează alţi roboţi. În „fabricile întunecate" – în care nu sunt oameni şi, prin urmare, nu este nevoie de lumină – roboţii lucrează zi şi noapte pentru construirea altor roboţi. Compania japoneză FANUC, unul dintre cei mai mari producători de roboţi industriali, deţine o fabrică întunecată în apropiere de muntele Fuji încă din anul 2001. Aceasta a ajutat FANUC să obţină vânzări anuale în jur de 6 miliarde de dolari, prin vânzarea de roboţi pe pieţele în creştere rapidă, cum ar fi China.

Un alt coleg a sugerat că vom ajunge să avem psihologi pentru roboţi. Chiar pare plauzibil să fie nevoie de câte un psiholog pentru fiecare robot? Consilierea psihologică a roboţilor va fi realizată de – cel mult – câţiva oameni în întreaga lume. Aşa că nu vor exista multe slujbe pentru îngrijirea sănătăţii mentale a roboţilor. Nu, va trebui ca meseriile noi să ţină de domenii în care oamenii excelează sau de domenii în care alegem intenţionat să nu folosim maşini.

Dar până în anul 2062 este posibil ca roboţii să devină supraumani, aşa că este greu să ne imaginăm o slujbă

în care oamenii le vor rămâne superiori. Asta înseamnă că profesiile rămase vor fi doar cele pe care preferăm să le aibă lucrătorii umani. Maşinile ar putea fi mai bune decât noi la aceste sarcini, atât din punct de vedere psihic, cât şi cognitiv, dar, cu toate astea, vom alege să fie făcute de oameni.

Roboţi nedoriţi

Atunci, revoluţia IA va fi despre redescoperirea lucrurilor care ne fac umani. Acesta este încă un motiv pentru care ar putea fi numită A Doua Renaştere. Ne vom redescoperi umanitatea.

Din perspectivă tehnică, maşinile vor deveni artişti extraordinari până în anul 2062. Vor fi capabile să scrie piese de teatru la fel de bune ca *Macbeth* a lui Shakespeare. Vor picta tablouri la fel de provocatoare ca *Guernica* lui Picasso. Şi vor compune muzică la fel de frumoasă precum cea a lui Erik Satie. Dar tot vom prefera opere produse de artişti umani. Aceste opere vor vorbi despre experienţa umană.

Vom aprecia un artist uman care va vorbi despre dragoste, pentru că este lucrul pe care îl avem cu toţii în comun. Nicio maşinărie nu va simţi dragostea aşa cum o simţim noi. Chiar dacă îşi vor dezvolta o conştiinţă şi emoţii proprii, nu vor experimenta niciodată dragostea

umană. Vom valoriza un artist uman care vorbeşte despre mortalitate, pentru că o avem în comun şi pe aceasta. Sau un artist uman care vorbeşte despre spiritul uman. Sau despre dreptate şi corectitudine. Sau despre oricare altă parte a experienţei *umane.*

Ca şi în cazul artei, vom aprecia şi artizanatul cu ochi noi. Putem chiar să vedem deja începuturile acestui fenomen în cultura hipsterilor. Vom aprecia din ce în ce mai mult lucrurile create de mâini *umane.* Bunurile produse în masă, de roboţi, vor ajunge să fie mai ieftine, în timp ce lucrurile făcute de mână vor fi mai rare şi vor avea o valoare din ce în ce mai mare.

Vor exista multe slujbe pentru artizani. Producători de bere. De brânză. De vin organic. Olari. Din nou, toate acestea rezonează cu experienţa noastră umană comună. Vom prefera bolul de lemn creat manual, în locul celui mai ieftin şi perfect, creat de maşini. Ne vom aduce aminte de povestea tâmplarului care ne-a spus cum a găsit lemnul din care a creat bolul, în timp ce se plimba cu câinele prin pădure. Şi ne vom aduce aminte că fiica tâmplarului se afla lângă el, învăţând despre acest meşteşug antic.

Ca animale sociale, vom aprecia din ce în ce mai mult interacţiunile sociale cu alţi oameni. *Baristi* vor ajunge să apese doar nişte butoane pe maşini de cafea controlate de calculatoare, care fac cafeaua perfectă

întotdeauna, dar noi tot vom sta la rând ca să ne facă un om cafeaua pentru a sta la taclale. Pentru a avea parte de un zâmbet şi de experienţa interacţiunii umane. Vom prefera în continuare ca un vânzător uman să ne ajute să ne alegem o rochie. Un doctor uman care să ne dea nişte veşti proaste legate de analizele noastre la sânge. Un barman uman care să ne toarne un pahar de whisky şi care să ne ofere nişte cuvinte de consolare. Un antrenor uman care să ne ajute să fim în formă. Şi un judecător uman care să dea sentinţe în sălile de judecată.

Cele mai importante trăsături umane în anul 2062 vor fi inteligenţa socială şi cea emoţională, precum şi abilităţile artistice şi meşteşugăreşti. Nu vor fi aptitudinile ŞTIM care în prezent par atât de importante pentru a avea un serviciu. Ironia este că viitorul nostru tehnologic nu va fi despre tehnologie, ci despre umanitatea noastră. Iar meseriile din viitor vor fi cele mai umane.

0005
Sfârşitul războiului

Există o altă profesie care cel mai probabil va dispărea prin automatizare până în anul 2062, iar eu şi mulţi alţii ne temem de asta. Acest lucru a început să se

întâmple deja. Are loc o cursă a înarmării pentru conceperea roboţilor care pot înlocui oamenii de pe terenul de luptă. Presei îi place să-i numească „roboţi ucigaşi”, dar termenul tehnic este „arme autonome letale”, sau AAL.

Problema cu faptul că sunt numiţi roboţi ucigaşi este că asta trimite la imaginea lui Terminator şi, implicit, la tehnologii la care se va ajunge doar peste foarte mult timp. Dar nu Terminatorii mă îngrijorează pe mine, şi nici pe miile de colegi de-ai mei care lucrează în domeniul IA. Tehnologiile mult mai simple mă preocupă, care vor apărea în cel mai bun caz (sau în cel mai rău caz) în mai puţin de un deceniu. Nu îmi este frică de inteligenţele artificiale deştepte, ci de inteligenţele artificiale proaste. Vom oferi unor maşini care nu sunt suficient de capabile dreptul de a lua decizii ce ţin de viaţă şi de moarte.

Să luăm ca exemplu o dronă Predator. Aceasta este o armă semiautonomă, care poate zbura de una singură în cea mai mare parte a timpului. Dar tot există un soldat, de obicei într-un container din Nevada, care o controlează. Şi cel mai important este că există un soldat care ia decizia de a lansa o rachetă. Dar nu este nevoie decât de un mic pas mai departe şi acel soldat va fi înlocuit de un calculator. Într-adevăr, acest lucru

este deja posibil din punct de vedere tehnic[48]. Iar odată ce vom construi arme autonome atât de simple, va fi o cursă a înarmării pentru dezvoltarea unor versiuni din ce în ce mai sofisticate.

Lumea va fi un loc mult mai urât dacă, după 20 de ani, armele autonome letale vor deveni ceva comun şi vor exista legi în privinţa lor. Această posibilitate va reprezenta o evoluţie groaznică în domeniul conflictului armat. Dar nu este ceva inevitabil. Avem dreptul să alegem dacă vrem să o luăm pe acest drum – şi vom lua această decizie în următorii câţiva ani.

Tentaţia roboţilor ucigaşi

Pentru armată, este evidentă atractivitatea armelor autonome. Cel mai slab punct al dronei Predator este legătura prin radio cu baza. Şi dronele au fost sabotate prin blocarea legăturii radio. Aşa că, dacă poţi face în aşa fel încât drona să poată zbura, urmări şi ţinti de una singură, ai o armă mult mai puternică.

O dronă autonomă în totalitate îţi permite şi să te debarasezi de mulţi piloţi de drone care sunt bine plătiţi. Forţele Aeriene ale Statelor Unite ar putea să îşi

[48] Nu trebuie să mă crezi pe mine pe cuvânt că o dronă complet autonomă este deja posibilă; ministrul apărării din Regatul Unit a declarat de asemenea că este posibilă în prezent.

schimbe numele în Forţele Dronelor Statelor Unite. Deja au mai mulţi piloţi de drone decât piloţi care să conducă orice tip de avion; până în anul 2062 nu doar că vor exista mai mulţi piloţi de drone decât de orice alt tip de avion, ci mai mulţi piloţi de drone decât toţi ceilalţi piloţi la un loc. Şi deşi nu îşi riscă viaţa în misiunile de luptă, piloţii de drone tot suferă de stres posttraumatic în egală măsură cu restul piloţilor forţelor aeriene.

Armele autonome oferă multe alte avantaje operaţionale. Nu trebuie să fie hrănite sau plătite. Pot să lupte tot timpul. Vor avea acurateţe şi reflexe la niveluri supraumane. Nu va trebui să fie evacuate niciodată de pe terenul de luptă. Vor respecta fiecare ordin cu stricteţe. Nu vor comite nicio atrocitate şi nu vor încălca dreptul umanitar internaţional[49]. Vor fi soldaţii, marinarii şi piloţii perfecţi.

Din punct de vedere strategic şi militar, armele autonome sunt un vis. Ele permit forţelor armate să îşi crească operaţiunile fără să depindă de limitele impuse de forţa de muncă. Un programator poate să controleze sute sau chiar mii de arme autonome. Războiul va

[49] Voi reveni asupra afirmaţiilor cum că armele autonome vor putea să urmeze dreptul umanitar internaţional şi nu vor comite atrocităţi. Spre deosebire de argumentele multora dintre cei care sunt în favoarea armelor autonome, voi susţine că aceste afirmaţii sunt îndoielnice.

ajunge să fie industrializat. Armele autonome vor spori numărul de opţiuni strategice. Îi vor scoate pe oameni în afara pericolului şi se vor ocupa de cele mai riscante misiuni. I se poate spune Război 4.0.

În luna septembrie 2017, Vladimir Puţin a declarat că oricine va ajunge să fie primul în domeniul IA va conduce întreaga lume. El a făcut predicţia că viitoarele războaie vor fi purtate de drone: când dronele unei tabere vor fi distruse de dronele inamice, nu vor avea de ales şi se vor preda. Dar sunt mai multe motive pentru care acest vis al forţelor militare se va transforma într-un coşmar până în anul 2062.

Moralitatea roboţilor ucigaşi

În primul şi-n primul rând, există un argument puternic din punct de vedere moral împotriva roboţilor ucigaşi. Renunţăm la o parte esenţială a umanităţii noastre dacă oferim unei maşini posibilitatea de a decide dacă cineva trebuie să moară sau să trăiască. Cu siguranţă în ziua de azi automatele nu au sentimente, compasiune sau empatie. Vor fi vreodată potrivite pentru a decide cine trăieşte şi cine nu?

Războiul este ceva îngrozitor. Se pierd vieţi. Oamenii sunt mutilaţi violent şi oribil. Civilii sunt bombardaţi. Populaţia este terorizată. În război acţionăm în feluri

care nu sunt permise în timp de pace. În parte permitem acest lucru deoarece soldaţii care comit aceste lucruri îşi pun propria viaţă în pericol. Îţi este permis să îţi omori inamicul întrucât, în acel moment, este viaţa ta sau a lui în joc.

Pentru că războiul este ceva îngrozitor, părerea mea este că nu ar trebui să fie uşor de făcut. Nu trebuie să fie ceva pentru care luptăm uşor şi „curat". Dacă istoria ne-a învăţat ceva, este că promisiunea unor războaie curate este şi va rămâne, cel mai probabil, doar o iluzie. Războiul trebuie să rămână întotdeauna o opţiune doar în ultimă instanţă. Politicienii trebuie să justifice de ce fiii şi fiicele noastre se întorc acasă în saci de cadavre.

Istoria tehnologiei militare ne arată că uciderea a ajuns să aibă loc de la distanţă din ce în ce mai mare. La început ne luptam direct, corp la corp. Praful de puşcă ne-a permis să ne dăm mai înapoi şi să luptăm de la o oarecare distanţă. Avioanele cu reacţie ne-au ajutat să atacăm de sus. Iar acum, tehnologii precum dronele ne permit să omorâm oameni de la mare distanţă, fără să ne mai riscăm vieţile. Armele autonome sunt pasul suprem pe care-l vom face pentru a ne elibera de actul războiului. Maşinile, şi nu oamenii, vor fi cele care vor omorî de unele singure, şi niciun om nu va fi implicat. Asta schimbă caracterul războiului în mod fundamental. Şi cu aceste schimbări, multe scuze

morale pe care le avem pentru război vor începe să se năruie.

Arme de distrugere în masă

Dincolo de argumentele morale, există mai multe aspecte tehnice şi legale din cauza cărora ar trebui să ne facem griji. Din punctul meu de vedere, unul dintre cele mai puternice argumente în favoarea interzicerii acestor arme este că vor revoluţiona războiul.

Prima revoluţie în cadrul războiului a avut loc odată cu inventarea prafului de puşcă de către chinezi. A doua a fost apariţia armelor nucleare, create de Statele Unite ale Americii. Fiecare a reprezentat o schimbare treptată a vitezei şi a eficienţei cu care am putut să omorâm. Armele autonome letale vor constitui cea de-a treia revoluţie.

Armele autonome vor fi arme de distrugere în masă. Înainte, dacă voiai să provoci daune, aveai nevoie de o armată de soldaţi. Trebuia să convingi acea armată să-ţi urmeze ordinele, dar trebuia să-i şi antrenezi, să-i hrăneşti şi să-i plăteşti. În viitor, un singur programator va putea să controleze sute sau chiar mii de arme. La fel ca în cazul oricărei alte arme de distrugere în masă – arme chimice, biologice şi nucleare – va fi necesar să

interzicem armele autonome[50].

În anumite privinţe, armele autonome letale sunt chiar şi mai periculoase decât cele nucleare. Ca să construieşti o bombă nucleară este necesar un grad de sofisticare tehnică foarte mare. Ai nevoie de resursele unui stat-naţiune şi de acces la materialele fisibile cu rol de combustibil nuclear. Ai nevoie de fizicieni şi de ingineri pricepuţi. Deoarece pentru fabricarea lor sunt necesare toate aceste resurse, armele nucleare nu s-au răspândit la scară largă. Pentru armele autonome nu va fi nevoie de resurse greu de obţinut. Pur şi simplu iei o dronă mică, o programezi printr-o reţea neuronală care va identifica, va urmări şi va lua în vizor orice faţă de caucazian. Un astfel de software de recunoaştere a feţei poate fi găsit pe multe telefoane mobile inteligente din ziua de azi. Apoi ataşezi câteva grame de explozibil la dronă. Punând la un loc câteva tehnologii existente, obţii o armă simplă, ieftină, dar letală.

Să duci un camion plin cu 10.000 de astfel de drone în New York City ar fi echivalentul unui atac precum

---

[50] Armele de distrugere în masă sunt considerate adesea a fi orice arme capabile să ucidă un număr mare de oameni la întâmplare. Dar armele autonome ar putea fi capabile să discrimineze şi să aleagă, fără să ucidă neapărat la întâmplare. Ar putea fi cele mai discriminatorii arme care au fost create vreodată. Cu toate astea, Departamentul Apărării al SUA defineşte armele de distrugere în masă ca fiind orice arme „capabile să producă distrugere la scară largă sau să provoace victime în masă". Nu precizează drept criteriu că astfel de arme trebuie să ucidă la întâmplare, fără discriminare.

atentatul de pe 11 septembrie 2001. Nici măcar nu ai nevoie ca armele pe care le foloseşti să fie foarte precise. Să presupunem că numai una din zece drone funcţionează – tot ai putea omorî o mie de oameni în doar câteva minute. Cu o acurateţe de 50%, ajungi la un număr de 5000 de morţi cât ai clipi[51].

Construirea unor astfel de arme este mult mai uşoară decât construirea unei maşini autonome. În cazul unui autovehicul, un procent de 99,99% fiabilitate s-ar putea să nu fie posibil, dar acelaşi procent ar fi mai mult decât suficient pentru o dronă ucigaşă. Având în vedere că mulţi producători de automobile plănuiesc să creeze maşini care să fie autonome în totalitate până în anul 2025, ne-am putea aştepta ca producătorii de arme să dezvolte şi ei astfel de drone ucigaşe în doar câţiva ani.

Arme ale terorii

Armele autonome de genul acesta vor fi arme ale terorii. Îţi poţi imagina cât de terifiant ar fi să fii urmărit de un roi de drone autonome? Ar putea pica în mâinile teroriştilor şi ale liderilor statelor rebele, oameni care

[51] Ca să punem lucrurile în perspectivă, software-ul de recunoaştere a imaginilor din ziua de azi reuşeşte să identifice corect figuri umane cu o rată mai bună decât o încercare din zece. Acurateţea actuală depăşeşte 19 identificări corecte din 20 de încercări.

nu au nicio remuşcare şi le vor folosi asupra cetăţenilor. AAL vor reprezenta armele ideale pentru reprimarea unei populaţii. Spre deosebire de oameni, armele autonome nu vor ezita când va veni vorba de comiterea unor atrocităţi, inclusiv genocid.

Există persoane care susţin că roboţii pot să fie mai morali decât soldaţii umani. Din punctul meu de vedere, acesta este cel mai interesant şi cel mai dificil argument în favoarea armelor autonome. Dar ignoră faptul că încă nu ştim cum să construim arme autonome care vor respecta dreptul umanitar internaţional. Regulile războiului îţi cer să ţinteşti atacatorii, nu civilii, să acţionezi într-un fel care este proporţional cu ameninţarea şi să recunoşti şi să respecţi atunci când un combatant se predă, sau atunci când e rănit şi nu mai poate lupta. Încă nu ştim cum să construim arme autonome care să poată face aceste distincţii.

Până în anul 2062 mă aştept să reuşim să construim roboţi etici. Vieţile noastre vor fi pline de aparate autonome, care vor trebui să acţioneze etic. Aşa că probabil vom avea şi AAL care vor respecta dreptul umanitar internaţional. Cu toate acestea, nu vom putea opri ca astfel de arme să fie atacate de hackeri şi forţate să acţioneze în feluri imorale. Dacă ai acces fizic la sistemul unui calculator, atunci e aproape sigur că poţi

intra în el. Şi sunt o mulţime de actori răi care vor putea ocoli mecanismele de siguranţă.

Ironia este că un număr de ţări – inclusiv Marea Britanie – se opun interzicerii armelor autonome letale tocmai pe baza argumentului că acestea încalcă oricum dreptul umanitar internaţional. Nu este nevoie de nicio legislaţie nouă în privinţa acestor arme, susţin ele. Istoria sugerează faptul că acest argument este greşit. Armele chimice încalcă legile umanitare internaţionale, în special *Protocolul de la Geneva* din anul 1925. Dar în 1993, *Convenţia privind armele chimice* a reglementat aceste arme mai strict. Convenţia a fost semnată şi ratificată de Marea Britanie, o ţară care acum susţine că dreptul umanitar internaţional actual e suficient şi pentru arme noi cum ar fi AAL.

Convenţia privind armele chimice a întărit legea internaţională pentru a interzice folosirea chimicalelor în război. A fondat Organizaţia pentru Interzicerea Armelor Chimice, o instituţie interguvernamentală cu sediul la Haga, pentru a monitoriza conceperea, producţia, stocarea şi folosirea armelor chimice. În prezent, peste 90% din stocurile de arme chimice declarate ale lumii au fost distruse. Interzicerea unor tipuri de arme poate avea un impact pozitiv asupra siguranţei şi a securităţii noastre.

Arme ale erorii

Pe lângă faptul că vor fi arme ale terorii, armele autonome vor fi şi arme ale erorii. Dintr-o perspectivă tehnică, ultimul loc în care ţi-ai dori să pui un robot este pe câmpul de luptă.

Există un motiv bine întemeiat pentru care roboţii au fost puşi pentru prima dată în locuri precum fabricile de maşini. Într-o fabrică, poţi să controlezi spaţiul. Poţi să decizi ce face fiecare. Poţi chiar să pui robotul într-o cuşcă, pentru a-i proteja pe spectatori. Câmpul de luptă este un mediu foarte diferit, plin de incertitudine şi de confuzie. Nu este locul în care ţi-ai dori să pui un robot care are potenţial letal.

În noiembrie 2016, o investigaţie condusă de *The Intercept* asupra operaţiunilor armatei americane împotriva talibanilor şi membrilor al-Qaida din munţii Hindukuş a arătat faptul că aproape nouă din zece oameni care au murit în urma atacurilor cu drone nu erau ţintele dorite. Reţineţi: asta se întâmplă în cazul în care avem un om implicat, fiind conştient de situaţie, şi care este superior oricărei maşinării. Şi acel om ia decizii de viaţă şi de moarte. Ca expert în tehnologie, dacă mi-aţi cere mie să înlocuiesc pilotul uman al dronei cu un robot, aş fi mulţumit dacă am putea să ajungem la nivelul performanţei umane şi am face doar

nouă din zece erori. Mă tem că am greşi aproape de fiecare dată.

Potenţialul de eroare este amplificat de viteza mult mai mare a acestor arme. Chiar dacă este implicat şi un om, maşinile tot s-ar putea să se mişte prea rapid pentru ca oamenii să intervină şi să prevină problemele. Şi sistemele unor astfel de arme se pot comporta în feluri neaşteptate. La fel ca la bursa de valori, s-ar putea să intre în nişte bucle de feedback neaşteptate. Dar spre deosebire de cazul bursei de valori, aici rezultatele vor fi mortale. Am putea ajunge chiar să purtăm războaie-fulger pe care nu am intenţionat să le purtăm. Erorile creează o problemă în plus: vidul de responsabilitate. Cine este responsabil atunci când armele autonome letale fac greşeli? Cine va fi chemat în faţa curţii marţiale? Cine va fi acuzat de crime de război la Haga? Lipsa unor responsabili este mare mai ales când arma foloseşte învăţarea mecanică pentru a învăţa cum să identifice şi să urmărească ţintele. În acest caz, producătorul nu programează de fapt comportamentul armei. Aceasta a învăţat de una singură. Şi ceea ce învaţă depinde de datele pe care le vede. Ba mai rău, armata va fi tentată să îi permită acestei arme să înveţe în continuare pe terenul de luptă. Dacă nu va face asta, un adversar va găsi rapid mijloace prin care să deruteze ceea ce este un program fix.

Pentru a preveni această situaţie, îşi va dori ca arma autonomă să se adapteze la orice va face inamicul, la fel cum ar face un soldat uman. Dar dacă o armă autonomă învaţă, un adversar determinat va căuta căi prin care să antreneze arma astfel încât să o neutralizeze. Ar putea chiar să o antreneze să îl atace pe cel care o controlează. Atunci cine va mai fi responsabil pentru erorile sale?

Stabilitate geopolitică

La nivel strategic, AAL reprezintă ameninţări care ar putea să destabilizeze înţelegeri temporare care există între ţări, precum cea dintre Coreea de Nord şi Coreea de Sud. Ar fi greu să lupţi în prezent împotriva unei armate de drone autonome mici şi viclene. Ameninţarea pe care o reprezintă ar putea tenta una dintre ţări să lanseze un atac-surpriză. Şi teama de un astfel de atac, precum şi incapacitatea de a se apăra de el ar putea elimina limitele folosirii de forţe şi mai mari, posibil chiar de arme nucleare.

Prin urmare, AAL ameninţă să perturbe actualul echilibru geopolitic. Orice ţară ar putea avea arme de distrugere în masă. AAL vor fi kalaşnicovurile viitorului. Spre deosebire de armele nucleare, vor fi ieftine şi uşor de produs. Şi vor apărea pe pieţele negre

din întreaga lume. Asta nu înseamnă că AAL nu pot fi interzise. Armele nucleare sunt ieftine şi uşor de produs, dar au fost interzise. Şi nu este nevoie să ne creăm propriile arme autonome ca să-i descurajăm pe cei care ar putea ignora o interdicţie, aşa cum a trebuit să facem în cazul armelor nucleare. Deja avem destui factori descurajând – militari, economici şi diplomatici – pe care să-i folosim împotriva celor care aleg să ignore tratatele internaţionale în privinţa armelor autonome letale.

Fiţi gata de luptă!

În luna iulie 2015, Max Tegmark, Stuart Russell şi cu mine am fost atât de alarmaţi de progresele din acest domeniu, încât ne-am adresat unui număr de o mie de colegi, cercetători care lucrează pe IA şi robotică, pentru a semna o scrisoare deschisă prin care solicitam ONU să interzică armele autonome ofensive. Scrisoarea a fost publicată la începutul celei mai mari conferinţe de IA, *Internaţional Joint Conference on Artificial Intelligence*[52]. O mie de semnături a părut a fi un număr frumuşel, şi reprezenta o proporţie substanţială din

[52] De ce conferinţa principală despre IA a fost numită International *Joint* („reunită") Conference on Artificial Intelligence este un lucru care s-a pierdut în istorie. Dar este totuşi locul în care poţi întâlni cercetători IA de pe tot globul.

comunitatea IA. Până să se încheie prima zi, numărul de semnături s-a dublat de la o mie la două mii. Şi a continuat să crească rapid pe parcursul conferinţei ce a ţinut o săptămână.

S-a vorbit mult în presă despre scrisoare, în special pentru că includea nume ale unor persoane foarte cunoscute, cum ar fi Stephen Hawking, Elon Musk şi Noam Chomsky. Dar din punctul meu de vedere, cel mai important a fost că era semnată de mulţi dintre cercetătorii de top în IA şi robotică. Proveneau din universităţi de pe tot globul, precum şi din laboratoare comerciale, cum ar fi Deepmind de la Google, AI Research Lab al Facebook şi Allen Institute for AI. Fără nicio discuţie, aceştia sunt oamenii care înţeleg cel mai bine tehnologiile şi limitele ei.

ONU a acordat atenţie avertismentului nostru. Scrisoarea a dus la iniţierea unor discuţii informale. Şi după doar un an, în decembrie 2016, în cadrul conferinţei principale de dezarmare, ONU a decis să continue şi cu discuţii formale pe această temă. AAL sunt acum luate în considerare de Grupul Experţilor Guvernamentali (GEG), o structură căreia Adunarea Generală a ONU i-a cerut să se ocupe de această problemă.

Dacă naţiunile pot ajunge la o înţelegere, eu sper că GEG va da o interdicţie în cadrul *Convenţiei pentru*

Anumite Arme Convenţionale. Titlul complet al acestei convenţii este de fapt: *Convenţia pentru interzicerea sau* restricţia folosirii anumitor arme convenţionale care pot fi excesiv de nocive sau care pot avea efecte aleatorii. Pentru diplomaţi, este cunoscută mai simplu sub denumirea de CAAC. Acesta este tratatul cu final deschis care a fost folosit înainte pentru a interzice terenurile minate, capcanele, armele incendiare sau laserele orbitoare.

Cursa înarmării

În scrisoarea deschisă, am avertizat că va exista o cursă a înarmării pentru dezvoltarea unor arme autonome din ce în ce mai capabile. Din păcate, această cursă a început deja. Pentagonul a alocat 18 miliarde de dolari din bugetul său pentru dezvoltarea unor tipuri noi de arme, multe dintre ele fiind autonome. Alte ţări, inclusiv Marea Britanie, Rusia, China şi Israel, au iniţiat şi ele programe complexe de dezvoltare a unor arme autonome.

Alege orice loc unde poate avea loc o bătălie – în aer, pe uscat, pe apă sau pe sub apă; există deja arme autonome pentru acel mediu la care armata lucrează în întreaga lume. Ai putea susţine că există cel puţin o armă autonomă care este deja funcţională. Aceasta este

SGR-A1 Sentry Guard Robot de la Samsung, care păzeşte zona demilitarizată (ZDM) dintre Coreea de Nord şi Coreea de Sud.

Nu ai niciun motiv pentru care să calci în ZDM. Este cea mai minată parte a lumii. Dar dacă nu te omoară minele, atunci te va omorî robotul de la Samsung. Poate să identifice automat, să ţintească şi să împuşte cu pistolul său autonom pe oricine calcă în zona-tampon.

Mai sunt şi alte arme la care se lucrează care ar putea fi considerate autonome. Putem exclude minele şi alte tehnologii simple, deoarece nu iau nicio decizie atunci când vine vorba de ţintire. Dar o armă precum sistemul Phalanx antirachete, care se află pe vapoare ale Marinei Regale Australiene, acţionează în mod autonom. Astfel, protejează baza de rachete supersonice cu ajutorul unui pistol controlat prin radar. Nu e suficient timp pentru ca oamenii să reacţioneze atunci când văd o rachetă la orizont. Sistemul antirachetă trebuie să identifice, să urmărească şi să ţintească autonom.

Există puţine motive de îngrijorare că există astfel de arme de apărare. Au un număr de operaţiuni foarte limitat. Ele protejează spaţiul aerian din jurul unei nave pe timp de război. Au ca ţintă doar obiecte care se mişcă cu o viteză supersonică. Mai degrabă salvează vieţi umane. Cei mai mulţi oameni, inclusiv eu, au

puţine obiecţii cu privire la utilizarea atât de limitată a autonomiei.

Pe de altă parte, o dronă autonomă care ar putea să hoinărească pe terenul de luptă timp de zile întregi ar constitui o problemă mult mai mare. Obiectivul operaţiunilor sale ar fi unul la o scară mult mai largă, atât din perspectiva timpului, cât şi a spaţiului. Dacă apare un convoi sub ea, trebuie să decidă singură dacă este un convoi militar, un convoi pentru ajutor sau o mulţime care participă la o nuntă. În prezent, nu prea te poţi baza pe aceste maşinării să facă astfel de distincţii.

La fel ca în cazul proliferării armelor nucleare, lumea este blocată într-o direcţie indezirabilă. Nu vrem o lume cu roboţi ucigaşi, dar dacă inamicii noştri au astfel de roboţi, am face bine să-i avem şi noi pe ai noştri – sau cel puţin aşa susţin cei care sunt împotriva unei interdicţii. Aşadar, o cursă a armelor a început deja să ducă la crearea unor arme pe care am prefera să nu le avem.

Şi adevărul este că s-ar putea să nici nu avem nevoie de arme autonome pentru a ne proteja împotriva unui atac cu arme autonome. Statele Unite, spre exemplu, explorează în prezent tehnologii mult mai simple, cum ar fi plasele şi păsările de pradă care să apere oamenii împotriva acestor drone controlate de la distanţă.

Obiecţii împotriva unei interdicţii

Au fost propuse mai multe argumente împotriva unei interdicţii a roboţilor ucigaşi. Din punctul meu de vedere, niciunul nu merită să fie examinat prea mult.

Una dintre cele mai serioase obiecţii împotriva unei interdicţii este aceea că roboţii se vor comporta mai etic decât soldaţii umani. Dar, aşa cum am explicat mai sus, încă nu ştim cum să construim roboţi etici. Şi nu ştim dacă IA va dobândi vreodată compasiunea şi empatia de care are nevoie pentru a se comporta în mod etic. Chiar dacă presupunem că am putea construi astfel de roboţi, nu ştim cum să facem în aşa fel încât să nu fie sabotaţi de hackeri şi puşi să se comporte imoral.

Un alt argument este că folosirea roboţilor ar implica faptul că soldaţii umani vor fi în afara oricărui pericol. Unii critici merg atât de departe, încât susţin că suntem astfel obligaţi din punct de vedere moral să folosim roboţi. Poate că cel mai tulburător aspect al acestei viziuni este că îi ignoră pe cei care vor fi nevoiţi să se confrunte cu roboţii ucigaşi. AAL vor creşte viteza cu care vom putea învinge tabăra adversă, şi astfel va creşte motivaţia pentru a porni un război. În cele din urmă, rezultatul ar fi mai mulţi morţi, nu mai puţini. Nu putem să ţinem cont numai de victimele din tabăra

căreia îi aparţinem.

O a treia obiecţie este că e imposibil să defineşti „armele autonome” – deci cum am putea interzice ceva ce nu poate nici măcar să fie definit? Sunt de acord, este dificil să oferim o definiţie a autonomiei. În domeniul IA suntem obişnuiţi cu asta. Majoritatea cercetătorilor au renunţat la încercarea de a defini ce este inteligenţa artificială; pur şi simplu ne ocupăm de construirea unor maşinării care să fie din ce în ce mai capabile. Mă aştept ca o eventuală interdicţie să nu dea o definiţie a „armei autonome”. Ar trasa doar o linie ce nu poate fi depăşită.

o dronă autonomă în totalitate care zăboveşte zile întregi deasupra câmpului de luptă ar fi probabil de partea interzisă a liniei. Dar acordul internaţional ar putea spune că o armă autonomă de apărare precum sistemul antirachetă Phalanx ar trebui să se afle pe lista lucrurilor neinterzise. Pe măsură ce apar noi tehnologii, vor apărea şi acorduri pentru legalizarea lor.

O a patra obiecţie este că noile tehnologii militare au făcut ca lumea să fie un loc mai sigur şi mai puţin violent, şi astfel ar trebui să primim cu braţele deschise armele autonome. Adesea se invocă astfel de argumente, ce au fost prezentate de Steven Pinker în

The Better Angels of Our Nature[53]. Pinker argumentează convingător că lumea din ziua de azi este locul cel mai puţin violent, care are cele mai puţine genocide din istoria omenirii.

Dar nimic din ceea ce spune Pinker nu contrazice necesitatea unei interdicţii. Impactul distrugător al noilor tehnologii a fost limitat doar de adoptarea dreptului umanitar internaţional şi a tratatelor noilor arme. Într-adevăr, bombele Austriei care au atacat Veneţia în anul 1849 – potrivit majorităţii surselor, a fost prima campanie de bombardare aeriană – au dus la *Convenţia de la Haga* din 1899, în cadrul căreia au fost interzise bombardările aeriene. Cât despre alte tehnologii noi, este nevoie de o lege nouă pentru a se limita folosirea roboţilor ucigaşi.

Al cincilea argument se referă la faptul că, spre deosebire de alte tehnologii care au fost deja interzise, cum ar fi laserele orbitoare, în cazul AAL vorbim despre o capacitate foarte generală, care ar putea fi adăugată aproape oricărei arme existente. Ar fi ca şi cum am încerca să interzicem folosirea curentului electric. Mai mult de atât, multe arme din prezent deja au o anumită formă limitată de autonomie, şi va fi imposibil de verificat dacă o armă semiautonomă a

---

[53] Vezi Steven Pinker, *The Better Angels of Our Nature*, Viking Books, New York, 2011.

avut o actualizare software pentru a o face complet autonomă.

Acest argument se bazează pe o înţelegere greşită a felului în care funcţionează tratatele despre arme. Nu există niciun regim de inspecţie pentru laserele orbitoare. Nu există nicio poliţie care să se asigure că nicio companie producătoare de arme nu construieşte mine. Dacă are loc vreo încălcare, ONG-uri cum ar fi Human Rights Watch se ocupă de documentarea ei. Apar ştiri peste tot în lume care condamnă faptele respective. Se adoptă rezoluţii de către Naţiunile Unite. Iar o condamnare de către tribunalul de la Haga rămâne un pericol real, deşi îndepărtat.

Toate acestea par să fie suficiente pentru a ne asigura că tratatele despre arme sunt încălcate rar. Asigură faptul că producătorii de arme nu vând arme interzise, că acestea nu se pot găsi pe piaţa neagră şi că nu ajung în mâinile teroriştilor. Am putea spera să obţinem ceva similar şi în cazul armelor autonome.

Paşi împleticiţi

Naţiunile Unite au decis la sfârşitul anului 2016 că GEG pentru arme autonome se va ţine de două ori în anul 2017, mai întâi în august, iar apoi în noiembrie, exact înainte de conferinţa anuală CAAC. Din păcate,

deşi diplomaţii au recunoscut că trebuie făcute progrese urgente în privinţa acestei situaţii, întâlnirea din luna august a fost anulată.

ONU a adoptat reguli noi de contabilitate şi fiecare ţară trebuie să plătească pentru fiecare întâlnire. Există câteva state, în special Brazilia, care au rămas în urmă cu contribuţiile. Din câte ştiu, Brazilia nu are neapărat ceva împotriva discuţiei despre roboţi ucigaşi. Doar că nu şi-a plătit datoriile de mai mulţi ani. Suma de bani de care a avut nevoie ONU pentru a ţine întâlnirea din august era de doar câteva sute de mii de dolari. Ăsta e nişte mărunţiş având în vedere că scopul este de a face lumea un loc mai bun şi mai sigur, aşa că am ajutat personal prin găsirea unui donator.

Dar, spre ruşinea Organizaţiei Naţiunilor Unite, a refuzat această donaţie caritabilă pentru plata întâlnirii din luna august. ONU ia bani numai de la guverne, au spus ei – ceea ce nu prea se potriveşte cu faptul că Ted Turner le-a oferit un miliard de dolari la sfârşitul anilor ’90. Drept consecinţă, din cauza a mai puţin de un sfert de milion de dolari, chestiunea a rămas nediscutată.

Ca să atrag atenţia asupra acestor întârzieri, am decis să acţionez la nivel public. La acea vreme, era o singură companie care declarase public că este împotriva armelor autonome: firma canadiană Clearpath Robotics. Aşa că am găsit fondatori de la o sută de

companii de robotică şi IA pentru a se semna o a doua scrisoare deschisă către CAAC pentru ca aceştia să facă ceva în privinţa roboţilor ucigaşi.

Am lansat scrisoarea din nou la deschiderea conferinţei internaţionale de IA, Internaţional Joint Conference on Artificial Inteligence, care s-a ţinut la Melbourne. Din întâmplare, conferinţa a început exact în ziua din august 2017 în care ar fi trebuit să aibă loc prima întâlnire a GEG. Această a doua scrisoare a fost semnată de Demis Hassabis şi de Mustafa Suleyman, doi dintre fondatorii Deepmind, precum şi de mulţi oameni bine cunoscuţi în domeniul IA şi al roboticii. Printre semnatari au fost şi Geoffrey Hinton şi Yoshua Bengio, doi dintre inventatorii învăţării profunde, şi Elon Musk, fondatorul companiei OpenAI.

La fel ca în cazul primei scrisori, şi aceasta a devenit virală în toată lumea. A demonstrat faptul că atât industria, cât şi mediul universitar susţin ideea reglementării acestor tehnologii. De asemenea, a introdus şi o frază ce a fost repetată foarte des de presă: „Odată ce această cutie a Pandorei va fi deschisă, va fi foarte greu de închis”.

La sfârşitul anului 2017, eu şi 137 de fondatori ai unor companii de IA şi de robotică ce am semnat scrisoarea am fost votaţi candidaţi la Omul Anului în competiţia anuală a Asociaţiei pentru Controlul Armelor, pentru

cea mai importantă contribuţie în favoarea dezarmării. Câştigătorii de drept au fost diplomaţii care au negociat cu ONU tratatul pentru interzicerea armelor nucleare. Dar a fost plăcut să văd faptul că problema armelor autonome a fost luată atât de în serios[54].

Până şi companiile producătoare de arme pot avea un beneficiu datorită acestei interdicţii. BAE Systems este unul dintre cei mai mari exportatori de arme, fiind o companie care lucrează cu prototipuri de sisteme autonome din generaţia următoare. La World Economic Forum din anul 2016, preşedintele companiei, John Carr, a susţinut că armele complet autonome nu vor putea să urmeze regulile războiului. Mai apoi a cerut guvernului să le reglementeze.

Presiune în creştere

Douăzeci şi trei de ţări au cerut până acum Naţiunilor Unite să interzică armele autonome letale. Acestea sunt: Algeria, Argentina, Austria, Bolivia, Brazilia, Chile, Cuba, Costa Rica, Ecuador, Egipt, Ghana, Guatemala, Sfântul Scaun, Irak, Mexic,

[54] Papa Francisc nu este probabil obişnuit să iasă pe locul trei, dar aşa s-a întâmplat în competiţia Asociaţiei pentru Controlul Armelor din anul 2017 pentru Omul Anului. Şi mă îndoiesc că este obişnuit să fie bătut de un profesor obscur din Australia.

Nicaragua, Pakistan, Panama, Peru, Palestina, Uganda, Venezuela şi Zimbabwe. În plus, Uniunea Africană a cerut să se dea o interdicţie preventivă. Mai recent, China a cerut să se interzică folosirea (dar nu şi dezvoltarea şi testarea) armelor complet autonome.

Va mai dura ceva timp până când susţinerea interdicţiei va deveni o opinie majoritară în cadrul Naţiunilor Unite, darămite un acord. Ţările care au sprijinit această interdicţie până acum sunt cele care ar putea fi atacate de nişte arme atât de teribile. Dar începe să se creadă din ce în ce mai mult că este necesar un „control uman semnificativ" al oricărui atac individual. Acest lucru i-ar cere tehnologiei să fie previzibilă, utilizatorului să aibă informaţii relevante, şi să existe potenţial pentru judecata şi intervenţia umană la timp.

Alte ţări încep să fie presate să acţioneze. În noiembrie 2017, exact înainte ca GEG despre armele autonome să aibă loc pentru prima dată la ONU, prim-ministrul Australiei a primit o scrisoare prin care se cerea Australiei să devină următoarea ţară care să semneze pentru interdicţia preventivă. Scrisoarea a fost semnată de peste o sută de cercetători din domeniul IA şi al roboticii din universităţile australiene. Din dorinţa de a se ajunge la un rezultat, eu am scris şi am organizat această scrisoare. Prim-ministrul Canadei a

primit o petiţie asemănătoare, semnată de peste două sute de cercetători canadieni; aceasta a fost organizată de colegul meu, profesorul Ian Kerr, care deţine titlul de preşedinte al Cercetărilor Canadiene în etică, drept şi tehnologie în cadrul Universităţii din Ottawa.

Scrisoarea australiană susţine că armele autonome letale care nu sunt controlate în niciun fel semnificativ de oameni depăşesc o limită etică clară. Cere guvernului nostru să îşi anunţe sprijinul în ceea ce priveşte interzicerea unor astfel de arme. „În acest fel, guvernul nostru îşi poate recăpăta poziţia de lider moral în lume, aşa cum a demonstrat anterior în domenii precum neproliferarea armelor nucleare", se spune în scrisoare. Scrisoarea canadiană exprimă lucruri similare.

Având în vedere că de curând Australia a fost aleasă în Consiliul pentru Drepturile Omului din cadrul ONU, problema armelor autonome este şi mai apăsătoare pentru Australia. Armele autonome reprezintă, în mod fundamental, o problemă ce ţine de drepturile omului. Raportorul special al Consiliului Organizaţiei Naţiunilor Unite pentru Drepturile Omului, profesorul Christof Heyns, a fost primul care a cerut Naţiunilor Unite să se ocupe de problema armelor autonome, susţinând în anul 2013 că maşinăriile nu ar trebui să aibă niciodată puteri asupra

vieţii oamenilor[55].

În ultimii ani, comunităţile din domeniile IA şi robotică au trimis un mesaj clar şi coerent despre armele autonome. Am avertizat că va exista o cursă a armelor, iar acum putem vedea că a început această cursă. De asemenea, am avertizat şi că vor exista riscuri considerabile din punct de vedere tehnic, legal şi moral în privinţa introducerii armelor autonome pe câmpul de luptă. Cât despre schimbările climatice, sunt câteva voci ştiinţifice în dezacord. Avem nevoie de un moratoriu, nu de o interdicţie, spun ei. Dar majoritatea avertizează că vor exista pericole mari şi cere să se dea o interdicţie preventivă.

Alternative la interdicţie

Marea Britanie este de părere că armele complet autonome încalcă dreptul umanitar internaţional care există în prezent, ea nu va dezvolta astfel de arme şi consideră că nu sunt necesare tratate noi pentru această chestiune. Există ceva adevăr în prima parte a declaraţiei lor. Dar nimeni nu ne garantează a doua parte. În trecut, chiar şi Anglia a creat arme chimice şi

---

[55] Vezi Christof Heyns, *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions*, Consiliului Organizaţiei Naţiunilor Unite pentru Drepturile Omului, 2013.

biologice în secret. Şi istoria contrazice a treia parte. Tehnologiile noi au făcut necesar ca legile umanitare internaţionale să fie întărite de-a lungul ultimului secol.

Alternativa principală pe care o propune Marea Britanie în locul interdicţiei este aşa-numitul Articolul 36 pentru verificarea armelor. Articolul 36 din Protocolul Adiţional I din cadrul *Convenţiei de la Geneva* cere statelor să verifice armele, mijloacele şi metodele de război noi, pentru a se asigura că respectă legile umanitare internaţionale. Anglia face astfel de verificări pentru orice sistem nou de arme.

Părerea mea este că verificările pe baza Articolului 36 nu sunt o alternativă bună, din mai multe motive. În primul rând, nu există nicio procedură standard acceptată pentru verificările de arme. Cum ne putem asigura că Rusia (ca să nu aleg o ţară chiar la întâmplare) verifică armele noi la fel de strict ca Anglia? În al doilea rând, nu avem niciun exemplu de sistem de arme care să nu fi trecut de verificare pe baza Articolului 36. Asta nu sugerează faptul că verificările reuşesc să ţină vreun tip de tehnologie departe de câmpul de luptă. Şi în al treilea rând, sunt doar câteva ţări care fac verificări potrivit Articolului 36 – cele care sunt obligate să îşi publice rezultatele.

Cum să evităm acest viitor

Ne aflăm la o răscruce în ceea ce priveşte această problemă. Putem alege să nu facem nimic şi să lăsăm companiile de arme să producă şi să vândă AAL. Asta ne va aduce într-un punct foarte periculos. Sau putem să intervenim şi să determinăm Naţiunile Unite să facă ceva în această privinţă.

Comunitatea de cercetare a transmis un mesaj clar. La fel şi industria. Şi din experienţa mea legată de acest subiect la nivel global, cea mai mare parte a publicului susţine cu tărie această interdicţie. Un sondaj realizat în anul 2017 de către compania Ipsos în 23 de ţări a descoperit că, în multe locuri, cei mai mulţi participanţi s-a opus folosirii de arme automate.

În cazul celor mai multe arme din trecut, a trebuit să fim martorii folosirii lor înainte să luăm atitudine. A fost nevoie să observăm efectele groaznice pe care le-au avut armele chimice în timpul Primului Război Mondial înainte să acţionăm şi să iniţiem Protocolul de la Geneva din anul 1925. A trebuit să fim martorii ororilor care s-au petrecut la Hiroshima şi la Nagasaki, şi să trăim experienţa celor câtorva momente din timpul Războiului Rece când era cât pe ce să izbucnească un conflict înainte să interzicem armele

nucleare[56]. Avem doar un singur caz – acela al laserelor orbitoare – în care a fost introdusă o interdicţie preventivă.

Cea mai mare temere a mea este că va trebui să fim martorii impactului terifiant al armelor letale autonome înainte să avem curajul de a le declara ilegale. Orice se va întâmpla, până în anul 2062 va fi văzut ca fiind ceva inacceptabil din perspectivă morală ca maşinile să decidă cine trăieşte şi cine moare. În acest fel, poate că vom putea să ne salvăm de acest destin groaznic.

0006
Sfârşitul valorilor umane

Armele autonome constituie un exemplu care ne arată felul în care schimbările tehnologice pot ameninţa valorile umane de bază care susţin societatea. Considerăm că oamenii au cu toţii nişte drepturi de bază, cum ar fi dreptul la libertatea de gândire, de conştiinţă şi de a avea o religie. Îi ajutăm pe cei bolnavi şi pe cei bătrâni. Ne aşteptăm ca fiecare persoană să

[56] Tratatul Naţiunilor Unite privind interzicerea armelor nucleare din anul 2017 a fost semnat de 58 de naţiuni, şi ratificat de 7 dintre acestea. Va intra în vigoare atunci când 50 de state îl vor ratifica. Niciunul dintre statele cu arme nucleare şi nici membrii NATO (cu excepţia Olandei) nu l-au semnat, aşa că încă nu am văzut impactul său asupra dezarmării nucleare.

aibă o şansă. Ca bărbaţii şi femeile să fie trataţi egal. Ca rasa şi religia să nu fie lucruri pe baza cărora luăm decizii în privinţa oamenilor.

În torentul continuu de ştiri proaste, este uşor să uităm că sunt multe lucruri bune care se întâmplă zilnic în lume. Există gesturi de mare bunăvoinţă, precum o familie care adoptă un copil orfan. Un doctor care se ocupă gratis de tratarea cataractei în ţările aflate în curs de dezvoltare. Un donator de rinichi care face posibil un transplant. Dar să nu uităm şi de gesturile de bunătate mărunte. Cineva care găteşte masa pentru un vecin în vârstă. O bancnotă de zece dolari pusă în cutia unei persoane fără adăpost. Un străin care te prinde de braţ atunci când te împiedici.

Această bunătate este ceea ce ne defineşte. Am devenit specia dominantă de pe planetă nu doar pentru că suntem cei mai deştepţi, ci pentru că am învăţat să cooperăm. Ne-am organizat în familii şi alte grupuri sociale, în comune şi oraşe, ţări. Efectele de reţea şi alte economii datorate producţiei la scară largă ne-au oferit avantaje imense în faţa altor specii de pe planetă.

Cooperând cu toţii, trăim în afara legilor evoluţiei ale lui Darwin. Nu mai este vorba despre supravieţuirea celor mai puternici. Într-adevăr, ne mândrim cu faptul că îi protejăm pe cei defavorizaţi. În anul 1900, unul din trei copii murea înainte să atingă vârsta de cinci ani.

Azi raportul este de mai puţin de un copil din douăzeci. Peste un miliard de oameni au fost ajutaţi să scape de sărăcie extremă din anul 1990 şi până acum. Speranţa de viaţă s-a dublat în ultimii trei sute de ani. Nu suntem perfecţi, dar adesea avem grijă unii de alţii.

Cu toate acestea, valorile pe care le avem în comun nu sunt garantate pentru totdeauna. Societatea este din ce în ce mai divizată. Mişcările naţionaliste şi separatiste sunt în creştere. Rasismul încă este un lucru des întâlnit. Multe valori umane de bază sunt sub ameninţare. Dar până în anul 2062, IA ar putea să înrăutăţească lucrurile mult mai mult. Până atunci, am putea să cedăm o mare parte a deciziilor care ne afectează vieţile roboţilor, care, fie că aşa au fost creaţi sau nu, nu vor avea valorile noastre umane.

Prejudecăţi automate

Dacă îi ceri lui Google Translate să îţi traducă „Ea este doctor” în turcă, iar apoi rezultatul obţinut să îl traducă înapoi în română, vei obţine: „El este doctor”. Pe de altă parte, dacă îi ceri lui Google Translate să îţi traducă „El este bonă” în turcă, şi apoi să traducă iar în română, va reieşi: „Ea este bonă”. Limba turcă nu are genuri gramaticale, aşa că atât „el”, cât şi „ea” vor fi traduse cu acelaşi cuvânt: „o”. Dar când traduci înapoi

în română, Google demonstrează că are nişte prejudecăţi de modă veche în privinţa doctorilor şi a bonelor. Aşadar, este de înţeles îngrijorarea că maşinile ar putea să continue să perpetueze multe dintre prejudecăţile peste care oamenii s-au chinuit din greu să treacă.

Poate te vei întreba dacă nu cumva am ales special aceste exemple, aşa că permite-mi să îţi mai ofer nişte date legate de Google Translate. Ţine minte: în turcă, „o” poate să însemne şi „el”, şi „ea”.

TURCĂ	ROMÂNĂ
o bir aşçi	este bucătar
o bir mühendis	el este inginer
o bir hemşire	ea este asistentă medicală
o bir asker	el este soldat
o bir öğretmen	ea este profesoară
o bir sekreter	ea este secretară
o bir sevgili	este un iubit
o sevildi	era iubit
o evli	el este căsătorit
o bekar	el este singur
o mutsuz	este nefericit
o mutlu	este fericit
o tembel	el este leneş
o çalişkan	el lucrează din greu

Aş fi putut foarte uşor să aleg alte servicii de traducere, cum ar fi Microsoft Translator. Şi aş fi putut alege alte limbi. Spre exemplu, când traduci din engleză în germană, Google Translate transformă *the kindergarden teacher* în femininul *die Kindergärtnerin*, dar *the teacher* în masculinul *der Lehrer.*

Motivul pentru există această abordare sexistă este că toate aceste aparate de tradus, la fel ca alţi algoritmi de învăţare mecanică, se bazează pe statistici. Iar aceste statistici sunt generate de pregătirea unui program în funcţie de o serie de texte ce conţin astfel stereotipuri de gen. Expresia *die Kindergärtnerin* apare mai frecvent decât echivalentul masculin *der Kindergärtner.* Aşadar, reflectă o preferinţă care există deja în textele scrise. Dar este o înclinaţie pe care majoritatea oamenilor nu ar dori să o vadă existând din oficiu în societatea noastră. Este una ce poate fi eliminată dacă nouă – sau mai degrabă lui Google şi celorlalte companii care se ocupă cu traducerea automată – ne-ar păsa suficient de mult.

Au fost identificate multe alte exemple de preferinţe algoritmice. Spre exemplu, un studiu din anul 2015, condus de CMU, a descoperit că Google oferea mai multe reclame pentru slujbe mai bine plătite unor

bărbaţi decât unor femei[57]. Spre deosebire de alţi comentatori, eu nu dau vina direct pe Google pentru această înclinaţie. Ar putea fi la fel de bine vina algoritmului sponsorului, ca şi a motorului de căutare Google. Dar nu contează neapărat cui aparţin algoritmii, cu siguranţă nu va ajuta la lupta împotriva discriminării faptul că femeile continuă să primească oferte de muncă mai slab plătite decât bărbaţii.

Companiile de tehnologie mari, cum ar fi Google şi Facebook, trebuie să-şi asume responsabilitatea în acest caz. Chiar dacă oferă multe servicii gratis, au datoria de a nu perpetua prejudecăţile. Companiile de tehnologie au promovat mitul cum că algoritmii nu reflectă preferinţe umane inconştiente – ci că oferă *în mod obiectiv* cel mai bun rezultat. Această minciună a permis giganţilor din domeniul tehnologiei să evite să-şi asume responsabilitatea pentru algoritmii lor şi astfel au fost scutiţi de efortul de a rezolva problemele.

Este adevărat, aşa cum a observat directorul de cercetare de la Google, Peter Norvig, că oamenii nu se pricep deloc să ia decizii. Economiştii care se ocupă de comportament au descoperit că suntem părtinitori,

---

[57] Vezi Amit Datta, Michael Carl Tschantz & Anupam Datta, „Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination", *Proceedings on Privacy Enhancing Technologies*, vol. 1, pp. 92-112, 2015.

plini de prejudecăţi şi că adesea ne comportăm iraţional. Dar dacă nu suntem atenţi, s-ar putea să construim maşinării care vor fi la fel de părtinitoare ca noi. De fapt, algoritmii de azi sunt de multe ori mai răi decât oamenii. Spre deosebire de oameni, mulţi algoritmi nu pot explica felul în care iau decizii. Sunt nişte cutii negre care pur şi simplu oferă răspunsuri. În cazul unui om, pot întotdeauna să întreb de ce a luat o anumită decizie. Dar cu IA din prezent, nu pot decât să accept răspunsul pe care mi-l dă.

Busola imorală

Până în anul 2062, prejudecăţile algoritmilor se vor răspândi extrem de mult dacă nu vom acţiona cu determinare pentru a le limita. Există mai multe exemple care demonstrează faptul că deja ne pun societatea la încercare. Haide să aruncăm o privire asupra unui caz din Statele Unite: un algoritm de învăţare mecanică creat de Northpointe, numit BUSOLA[58]. Acesta a fost antrenat cu date din trecut pentru a oferi o predicţie cu privire la probabilitatea ca un infractor condamnat să recidiveze.

Acum, ai putea folosi un astfel de algoritm de

[58] COMPAS *(Correctional Offender Management Profiling for Alternative Sanctions*), în lb. eng. în orig. (n. red.).

învăţare mecanică în cadrul serviciilor de probaţiune, pentru a-i ajuta pe oamenii cei mai vulnerabili să nu ajungă din nou în închisoare. Mă îndoiesc că vor fi puţini cei care nu vor fi de acord cu acest mod de a folosi tehnologia. Ar putea transforma societatea într-un loc mai bun şi mai sigur. Dar BUSOLA nu este folosită astfel. Judecătorii utilizează programul pentru a decide ce sentinţă, ce cauţiune sau ce perioadă de eliberare condiţionată să stabilească.

Nu mai este nevoie să spun că acest lucru este mult mai problematic. Chiar poate un program să ia acelaşi gen de decizii precum un judecător experimentat? Poate să ia în considerare toţi factorii de care ţine cont un judecător atunci când decide pedeapsa cuiva?

Să presupunem pentru o clipă că până în 2062 vom avea un program computerizat care va lua în calcul toţi aceşti factori, şi că programul este de fapt mai precis decât un judecător uman. Am putea justifica faptul că judecătorii umani vor continua să dea sentinţe? Nu suntem obligaţi din perspectivă morală să lăsăm aceste decizii pe mâna unei maşinării superioare?

Problema nu e pur ipotetică. Un studiu din anul 2016, condus de firma de investigaţii Propublica, a descoperit că programul BUSOLA prezice că acuzaţii negri vor recidiva mai frecvent decât este cazul. În acelaşi timp, a prezis că cei albi vor recidiva mai puţin

frecvent decât se întâmplă de fapt[59]. Aşa că cel mai probabil oamenii de culoare sunt închişi pe nedrept pe o perioadă mai lungă în comparaţie cu cei albi din cauza prejudecăţilor unui program. Iar oamenii albi care vor comite din nou delicte sunt lăsaţi să umble liberi în societate. Mă îndoiesc că programatorii de la Northpointe au intenţionat ca programul BUSOLA să aibă preferinţe rasiale. Dar are. Nu ştim sigur de ce este aşa. Din motive comerciale, Northpointe a refuzat să ofere detalii legate de cum funcţionează BUSOLA. Secretomania de acest fel este şi ea problematică. Dar ştim că programul a fost antrenat după date istorice şi că datele istorice sunt cel mai probabil rasiste. Rasa nu a făcut parte din datele de intrare, dar codul poştal da. Şi în multe locuri acesta este un bun indicator al rasei. Poate că sunt mai multe patrule de poliţişti în cartierele unde locuiesc preponderent persoane de culoare, aşa că este mai probabil ca oamenii negri să fie prinşi comiţând un delict? Poate că sărăcia duce la multe crime, şi din moment ce unele coduri poştale aparţin unor zone sărace, de fapt penalizăm sărăcia?

Odată ce am identificat o prejudecată automată ca aceasta, putem încerca să o eliminăm. Ar trebui să decidem ce înseamnă pentru predicţiile unui program

---

[59] Vezi Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, „Machine Bias“, *ProPublica*, 23 mai 2016.

de învăţare mecanică să fie imparţiale din punct de vedere rasial, iar apoi să ne asigurăm că este antrenat astfel încât să evite astfel de părtiniri. Totuşi, chiar dacă facem asta, tot este îndoielnic faptul că ar trebui să lăsăm aceste decizii pe mâna programelor de computer. Privarea oamenilor de libertate este unul dintre cele mai dificile lucruri pe care trebuie să le decidem ca societate. Nu ar trebui să privim situaţia cu superficialitate. Renunţăm la o parte importantă a umanităţii noastre atunci când lăsăm roboţii să ia astfel de decizii.

În ciuda publicităţii negative considerabile de care a avut parte BUSOLA, se pare că erorile sale se vor repeta. În anul 2017, poliţia din nord-estul Angliei a început să folosească învăţarea mecanică pentru a-i ajuta pe oamenii legii să decidă dacă să ţină suspecţii în custodie sau nu. Instrumentul de Evaluare a Riscurilor (*HART*) foloseşte date din evidenţele poliţiei, precum şi din cazierul suspectului, şi anumite date demografice pentru a analiza cât de probabil este ca suspecţii să comită o infracţiune dacă vor fi eliberaţi. Din nou, codul poştal este unul dintre criteriile multiple utilizate pentru a formula o predicţie.

Algoritmii sunt folosiţi şi într-o serie de alte domenii asemănătoare. Curtea de probaţiune din Pennsylvania foloseşte predicţii bazate pe algoritmi de învăţare

mecanică din anul 2010 pentru a lua decizii legate de eliberarea condiţionată. Poliţia Metropolitană din Londra foloseşte un software dezvoltat de compania Accenture pentru a preconiza care membri ai unei bande vor comite din nou infracţiuni violente. Iar departamentele de poliţie din mai multe state americane, inclusiv California, Washington, Carolina de Sud, Arizona, Tennessee şi Illinois, folosesc un software care prezice unde şi când este cel mai probabil ca o infracţiune să se petreacă. În toate aceste cazuri nu a verificat nimeni dacă programele folosite sunt părtinitoare.

Discriminare algoritmica

Aşa cum am văzut, un motiv pentru care algoritmii pot să ia decizii părtinitoare este faptul că sunt antrenaţi cu date părtinitoare. Programul BUSOLA a fost antrenat să prezică cine va comite din nou un delict, dar nu a fost antrenat cu date despre prizonierii care chiar au recidivat. Nu ştim cine va recidiva. Unii o vor face, dar nu vor fi prinşi; ştim doar de cei care au fost arestaţi şi condamnaţi. Aşadar, datele ar putea să conţină anumite preferinţe rasiale şi de alte feluri, aspect ce se va reflecta în predicţiile programului.

Joy Buolamwini, cercetător la MIT Media Lab, a

fondat Liga Justiţiei Algoritmice pentru a descoperi astfel de prejudecăţi la software-ul care ia decizii. Fiind ea însăşi de culoare, a constatat că algoritmii de recunoaştere se chinuiau să o identifice; a folosit chiar şi o mască albă pentru a fi recunoscută. A descoperit că datele părtinitoare reprezintă sursa acestei probleme.

În cadrul comunităţii de recunoaştere facială există nişte seturi de date standard în funcţie de care poţi compara performanţele diverşilor algoritmi. Se presupune că, dacă un algoritm se descurcă bine cu aceste date de reper, atunci se va descurca bine per total. Dar nu ne-am pus problema cât de reprezentative sunt datele de reper, aşa că atunci când construim algoritmi care se descurcă bine cu datele standard ne creăm falsa impresie că am avansat.

Acum că analizăm acest lucru, pare foarte evident, dar atunci când lucrezi într-un laborator de cercetare e de înţeles faptul că faci un test rapid folosind orice şi pe oricine e prin preajmă – pui totul la un loc repede, ai un termen-limită, înţeleg de ce au apărut aceste probleme. Colectarea datelor, în principal a celor diverse, nu este un lucru uşor de făcut.[60]

---

[60] Ian Tucker, „A White Masked Worked Better: Why Algorithms Are Not Colour Blind“, *The Observer*, 28 mai 2017.

Unul dintre cele mai importante standarde în recunoaşterea facială este baza de date „Labelled Faces in the Wild”. Aceasta a fost publicată în anul 2007 şi conţine peste 13.000 de imagini cu feţe adunate din ştirile de pe internet. Dacă ne gândim la perioada în care a fost publicată, cea mai comună faţă era cea a lui George W. Bush. Setul de date conţine 77,5% bărbaţi şi 83,5% persoane albe. Este evident că persoanele care apar la ştiri nu sunt reprezentative pentru populaţia lumii.

Dar există seturi de imagini care se folosesc în cadrul comunităţii din domeniul viziunii computerizate, care sunt mai variate. Spre exemplu, „10k US Adult Faces Database”, publicată în anul 2013, conţine 10.168 de feţe, şi a fost creată special pentru a se potrivi cu distribuţia demografică din Statele Unite (potrivit unor variabile precum vârsta, rasa şi sexul). Iar Facebook are miliarde de poze la dispoziţie pentru cercetarea sa Deepface: de vreme ce aproape toată lumea care îşi face cont de Facebook încarcă o fotografie. Facebook chiar este o Carte de Feţe extrem de mare. Aşa că nu este deloc clar că recunoaşterea facială este ţinută pe loc de absenţa unor seturi de date de antrenament diversificate.

Mai este un factor simplu care ar putea face ca acest comportament părtinitor al roboţilor să continue, factor

care s-ar putea dovedi a fi mai problematic pentru mulţi liberali cu intenţii bune. În cazul oamenilor, există dovezi care atestă faptul că ne descurcăm mult mai bine să recunoaştem oameni din propriul nostru grup etnic decât din afara lui. Acest lucru se numeşte „efectul de rasă diferită”. Există efecte similare în cadrul şi între diferite grupuri de vârstă. Aşa că este posibil ca programul de recunoaştere facială să reproducă acest efect. O soluţie ar fi antrenarea unor algoritmi de recunoaştere facială separaţi pentru diverse grupuri de rase, şi pentru diverse grupuri de vârstă.

Există un fenomen asemănător şi în recunoaşterea vocii. Pentru a dobândi o acurateţe bună atât pentru vocile de bărbaţi, cât şi pentru cele de femei, ai nevoie de un software diferit. Aşa că preferinţele rasiale ale software-ului de recunoaştere facială s-ar putea să nu se datoreze datelor părtinitoare, ci pur şi simplu pentru că trebuie să folosim programe diferite de recunoaştere a feţelor.

Război de gorilă

Având în vedere că recunoaşterea facială se ocupă doar de feţe, probabil că nu este surprinzător faptul că software-ul pentru recunoaştere facială este predispus

la rasism. În anul 2015, Jacky Alciné a descoperit că Google Photos eticheta automat pozele cu el şi cu iubita lui ca „gorile”. Mesajul lui pe Twitter a descris problema succint:

> Google Photos, aţi dat-o-n bară.
> Iubita mea nu este o gorilă.

Din moment ce problema nu putea fi soluţionată uşor, Google a şters pur şi simplu eticheta „gorila”. Mulţi oameni au sugerat că de vină fusese baza de date. Nu ştim ce date a folosit Google Photos pentru a se antrena, dar problema ar putea fi că programele IA, şi în special reţelele neuronale, sunt instabile şi se pot da rateuri în feluri în care oamenii nu o fac.

Google Photos etichetează uneori oameni albi ca fiind „foci”. Dar acest lucru nu este considerat a fi la fel de jignitor ca etichetarea oamenilor negri ca fiind „gorile”. Dacă noi doi punem o etichetă unei poze, înţelegem faptul că, dacă punem greşit eticheta de gorilă unei persoane de culoare, asta va fi foarte ofensator. Dar programele IA nu au astfel de cunoştinţe banale. Noi considerăm că o astfel de etichetare constituie rasism. Dar ele nu au raţiune. Şi nici nu au habar ce înseamnă o ofensă.

Această situaţie pune în lumină diferenţele

fundamentale dintre inteligenţa umană şi inteligenţa artificială. Ca oameni, performanţa noastră în îndeplinirea sarcinilor tinde să scadă treptat pe măsură ce sarcinile se modifică. Dar sistemele IA se strică adesea în feluri catastrofale. Este un lucru important de ţinut minte atunci când oferim maşinilor puterea de a lua decizii. În special pentru că vieţile noastre sunt în joc, trebuie să fim foarte conştienţi de faptul că sistemele IA pot să dea rateu în feluri diferite în comparaţie cu oamenii, şi adesea mult mai dramatic.

Părtinire intenţionată

Sunt multe exemple de algoritmi care au fost creaţi în mod intenţionat pentru a fi părtinitori. În anul 2012, *Wall Street Journal* a descoperit că ziarul de călătorii *Orbitz* sugera utilizatorilor de Mac hoteluri mai scumpe decât celor care foloseau Windows[61]. Cei de la *Orbitz* au declarat că ei nu arată aceeaşi cameră la doi utilizatori diferiţi la preţuri diferite, dar nu avem decât cuvântul lor care să ateste acest lucru. În orice caz, este cel mai probabil ca *Orbitz* să sugereze unui utilizator de Mac o cameră considerată preferabilă sau o suită, şi o cameră obişnuită pentru un utilizator Windows.

---

[61] Dana Mattioli, „On *Orbitz*, Mac Users Steered to Pricier Hotels”, *The Wall Street Journal,* 23 august 2013.

Nu există nimic care să oprească preţurile dinamice să devină o obişnuinţă şi ca utilizatori diferiţi să primească preţuri diferite pentru aceeaşi cameră de hotel. Au fost multe cazuri în care Hertz mi-a sugerat o ofertă mult mai scumpă ca „membru Gold" decât atunci când am încercat să rezerv aceeaşi maşină ca utilizator de rând. În mare, suntem obişnuiţi ca toată lumea să aibă acces la preţuri similare. Percepem ca incorect faptul că unele grupuri sunt taxate mai mult pentru un bun sau un serviciu identic.

Preţurile dinamice pot părea incorecte, dar sunt legale în majoritatea ţărilor, atât timp cât nu sunt stabilite în funcţie de rasă, religie, naţionalitate sau sex şi atât timp cât legile locale nu sunt încălcate. Pieţele online oferă vânzătorilor mai multe oportunităţi pentru preţuri dinamice. Iar prin identificarea unor caracteristici care ne dezvăluie sensibilitatea la preţ – cum ar fi sistemul de operare pe care îl folosim – vânzătorii online îşi pot mări profitul.

Dar nu trebuie să ne resemnăm cu această situaţie. Am putea pur şi simplu să solicităm ca toţi consumatorii online să aibă acces la aceleaşi preţuri. Există unele pieţe unde deja limităm tarifarea discriminatorie. De exemplu, în anul 2012 Curtea Europeană de Justiţie a dat o sentinţă care spune că companiile de asigurări nu pot cere preţuri diferite

bărbaţilor şi femeilor. Drept rezultat, preţurile asigurărilor auto, de sănătate şi de viaţă din ţările din Uniunea Europeană nu mai depind de sexul persoanei asigurate.

Ironia este că există argumente rezonabile pentru a permite ca preţurile asigurărilor să discrimineze în funcţie de sex. Femeile tind să fie şoferi mai precauţi decât bărbaţii, şi femeile tind să trăiască mai mult decât bărbaţii. Asta înseamnă că companiile de asigurări pe viaţă şi pentru autoturisme trebuie să cheltuiască mai mult pentru bărbaţi decât pentru femei. În mod raţional, ar trebui ca ei să fie ei taxaţi mai mult. De ce să subvenţioneze femeile şofatul periculos al bărbaţilor? Asta nu face decât să încurajeze bărbaţii să continue să conducă într-un mod periculos.

Poate că este mai puţin rezonabil să îi taxezi pe bărbaţi mai mult pentru asigurarea de viaţă decât pe femei. Bărbaţii nu îşi aleg sexul. Pe de altă parte, comportamentul şi stilul de viaţă, precum şi genetica masculină îi omoară pe bărbaţi mai devreme. Este chiar mai discutabil să discriminăm pe baza sistemului de operare pe care îl foloseşte cineva. Poate că ar trebui să existe o legislaţie pentru astfel de tarifari?

În cele din urmă, dacă tarifarea discriminatorie ar fi dusă la extrem, asta ar face inutil să mai cumperi asigurări. Scopul asigurării este de a proteja indivizii

prin distribuirea riscului în cadrul unei populaţii. Tarifarea discriminatorie pune acest risc la loc pe umerii individului. Ca societate, acceptăm preţuri puţin mai mari pentru majoritatea, astfel încât să-i protejăm pe cei defavorizaţi. Ar trebui să lăsăm tehnologia să distrugă această solidaritate?

Părtinire ilegală

Au ieşit la iveală şi exemple de algoritmi intenţionat părtinitori în mod ilegal. De exemplu, în anul 2015 a fost descoperit că Volkswagen a folosit un algoritm de software sofisticat pentru câteva dintre maşinile sale diesel, care declanşa controlul deplin al emisiilor doar în timpul procedurilor de testare oficiale. În acest mod, nivelul măsurat oficial al emisiilor de oxid de azot a fost influenţat astfel încât motoarele să pară mai puţin poluante decât erau de fapt. Ulterior, Volkswagen a avut de plătit amenzi şi penalizări în valoare de 30 de miliarde de dolari.

Un alt exemplu a ieşit la lumină în anul 2017, când s-a descoperit că Uber îşi folosea ilegal software-ul „Greyball” pentru a contracara reglementările legale. Prin crearea unor geoperimetre în jurul clădirilor guvernamentale, prin folosirea informaţiilor despre cardurile de credit asociate cu funcţionari

guvernamentali şi a reţelelor de socializare, Uber a încercat să împiedice ca aplicaţia sa să fie folosită de membrii guvernului.

Faptul că ambele exemple implică maşini ar trebui să ne dea de gândit. În cursa pentru revoluţionarea transporturilor prin dezvoltarea vehiculelor autonome, a maşinilor şi a camioanelor electrice, precum şi a noilor mijloace de transport, ne putem aştepta ca multe alte companii de transport să fie tentate să se folosească de algoritmi în mod ilegal. Sunt în joc trilioane de dolari, aşa că recompensa potenţială este imensă. Şi avem foarte puţine mecanisme de protecţie care să prevină acest lucru.

Ce este corect?

Dacă vrem să cedăm anumite decizii către algoritmi, trebuie să fim mai expliciţi în ceea ce priveşte ce înseamnă să fii corect. Calculatoarele interpretează instrucţiuni într-un mod literal – frustrant de literal. Aşa că, dacă vrem să învăţăm calculatoarele să se comporte corect, trebuie să fim foarte precişi când vine vorba de ce înseamnă de fapt corectitudinea.

Haideţi să revenim asupra programului BUSOLA – dar ca să fac ca situaţia să fie mai puţin sensibilă din punct de vedere rasial, să ne gândim ce înseamnă

pentru un program de genul acesta să fie corect faţă de femei şi faţă de bărbaţi. Dreptatea ar trebui să fie oarbă şi, prin urmare, orice program precum acesta ar trebui să fie orb în privinţa sexului unei persoane. Dar ce înseamnă asta de fapt?

Există câteva tipuri diferite de corectitudine pe care am putea încerca să le codăm în programarea noastră. Un posibil criteriu simplu al corectitudinii ar fi ca procentul de bărbaţi despre care programul prezice că vor recidiva să fie egal cu procentul de femei despre care va spune acelaşi lucru. Dar ar fi simplist să procedăm astfel. Femeile poate că vor comite mai puţine crime decât bărbaţii, ceea ce ar duce la faptul că am închide femei nevinovate sau am elibera bărbaţi periculoşi doar ca să ne asigurăm că procentul este egal.

Un criteriu mai bun al corectitudinii ar fi ca, în general, acurateţea programului să fie identică atât pentru femei, cât şi pentru bărbaţi. Adică procentul de oameni care au fost clasificaţi incorect ar trebui să fie acelaşi pentru bărbaţi şi femei. Femeile ar fi furioase pe bună dreptate dacă un procent mai mare de femei ar fi clasificate greşit, mai ales dacă mai multe ar fi clasificate în mod greşit ca având un risc mare de recidivă.

Problema cu acurateţea generală folosită drept

criteriu pentru corectitudine este că prezicerea corectă a persoanelor care vor recidiva este mai importantă decât prezicerea corectă a persoanelor care nu vor recidiva. Eliberarea oamenilor care vor recidiva are costuri mai mari decât neeliberarea celor care nu vor recidiva. Punând cele două grupuri laolaltă, presupunem că sunt la fel de importante. În plus, nu respectăm proporţia lui Blackstone, potrivit căreia e preferabil ca guvernele şi instanţele judecătoreşti să greşească declarând pe cineva vinovat drept nevinovat, mai degrabă decât invers. Sir William Blackstone, în lucrarea lui influentă din secolul al XVII-lea, *Commentaries in the Laws of England*, a afirmat: „Este mai bine să scape zece persoane vinovate, decât să sufere o singură persoană nevinovată".

Un al treilea criteriu de corectitudine face distincţia dintre aceste două grupuri: cei vinovaţi declaraţi nevinovaţi şi cei nevinovaţi declaraţi vinovaţi. Am putea cere ca rata rezultatelor fals pozitive şi a celor fals negative să fie aceeaşi pentru bărbaţi şi pentru femei. Rata rezultatelor fals negative reprezintă câţi oameni au recidivat dintre cei despre care s-a prezis că nu o vor face raportat la numărul celor care au recidivat. Iar rata rezultatelor fals pozitive reprezintă raportul dintre câţi nu au recidivat dintre cei despre care s-a prezis că o vor face şi numărul celor care nu au recidivat. Femeile ar fi

ofensate pe bună dreptate dacă rata rezultatelor fals pozitive pentru femei ar fi mai mare decât cea pentru bărbaţi, şi prin urmare, mai multe femei ar fi trimise la închisoare pe nedrept. Din nou, falsele pozitive ar putea dăuna mai mult societăţii decât falsele negative. Aşadar, poate că ne vom dori să considerăm mai relevantă reducerea ratei rezultatelor fals negative.

Un al patrulea posibil criteriu de corectitudine împarte plăcinta predicţiilor cu totul altfel. Am putea cere ca ratele de eroare ale predicţiilor de eşec (cei care au recidivat de fapt) sau de succes (cei care nu au recidivat de fapt) să fie aceleaşi pentru bărbaţi şi pentru femei. Rata de eroare a predicţiei de eşec reprezintă raportul dintre numărul celor despre care s-a prezis că nu vor recidiva, dar au recidivat (rezultatele fals negative) şi numărul predicţiilor negative. Ceea ce e totuna cu a spune că rata de eroare a predicţiilor de eşec este numărul celor despre care algoritmul a prezis că nu vor recidiva şi care au făcut-o, împărţit la numărul celor despre care algoritmul a prezis că nu vor recidiva. Prin contrast, ne amintim ceea ce spuneam mai sus, rata falselor negative împărţea numărul de oameni despre care algoritmul a prezis că vor recidiva, dar nu au făcut-o, la numărul celor care au recidivat de fapt.

În mod similar, rata de eroare a predicţiei de succes

este raportul dintre numărul celor despre care s-a prezis că vor recidiva, dar nu au făcut-o (rezultatele fals pozitive) şi numărul predicţiilor pozitive. Ceea ce e totuna cu a spune că rata de eroare a predicţiilor de succes este numărul celor despre care s-a prezis că vor recidiva, dar nu au făcut-o, împărţit la numărul celor despre care algoritmul a prezis că vor recidiva. Prin contrast, ne amintim, rata falselor pozitive împărţea numărul celor despre care algoritmul a prezis că vor recidiva, dar nu au făcut-o, la numărul celor care nu au recidivat de fapt. Femeile ar fi din nou pe drept ofensate dacă rata erorii de predicţie de succes ar fi mai mare decât în cazul bărbaţilor, şi dacă acest lucru ar avea drept consecinţă închiderea femeilor pe nedrept.

Din nou, rata erorii de predicţie de eşec ar putea fi mai importantă pentru societate decât rata erorii de predicţie de succes, întrucât eliberarea oamenilor în societate care vor recidiva ar fi mai costisitoare decât păstrarea în închisoare a oamenilor care nu vor recidiva. Cei care sunt închişi pe nedrept s-ar putea să nu fie de acord, dar trebuie să ajungem la un echilibru cumva între libertatea lor şi siguranţa societăţii. Aşadar, va trebui să tratăm cele două rate de eroare diferit.

Am putea folosi alte criterii de corectitudine, cum ar fi raportul dintre rezultatele fals pozitive şi cele fals

negative sau tratamentul egal pentru indivizii „similari”. Dar acest lucru nu este important pentru discuţia noastră. Mesajul cu care trebuie să rămânem este acesta: corectitudinea poate însemna multe lucruri diferite. Nu există o definiţie unică şi simplă. De fapt, adesea se discută încontinuu în societate despre ce fel de corectitudine ar trebui să fie aplicată într-un anumit context. Şi acolo unde nu există acest tip de discuţie, ar trebui adesea să existe. Dacă vrem să încredinţăm luarea deciziilor în privinţa acestor chestiuni roboţilor, trebuie să ne gândim cu atenţie la ce vrem să însemne corectitudinea într-un anume context.

Transparenţă

Transparenţa în privinţa deciziilor maşinilor este de asemenea o chestiune vitală. Vrem ca maşinile să ia decizii corecte, dar să fie şi transparente atunci când fac acest lucru, pentru a ne încredinţa că sunt corecte. Asta este o provocare majoră pentru sistemele IA din prezent. Abordările la modă cum ar fi învăţarea profundă generează sisteme care nu îşi pot explica deciziile într-un fel care să aibă sens pentru noi. Deciziile sunt adesea produsul antrenamentului prin expunerea la mai multe date decât ar putea un om să proceseze într-o viaţă.

Bineînţeles, oamenii nu sunt nici ei prea transparenţi în general. Şi e notoriu faptul că „inventăm” explicaţii post-factum pentru deciziile noastre. Dar există o diferenţă fundamentală. Pe oameni putem să-i tragem la răspundere pentru deciziile lor.

Dacă decizia mea este extrem de proastă şi duce la moartea unui om, o să fiu acuzat de omor. Maşinăriile nu pot fi trase la răspundere astfel. Aşadar, este mai important ca acestea să fie capabile să explice de ce iau deciziile pe care le iau.

Transparenţa ne va ajuta să avem încredere în sistemele automate. Dacă o aplicaţie medicală îţi recomandă un tratament cu chimioterapie periculos, cei mai mulţi dintre noi am prefera un sistem transparent, care să ne explice cum a decis că avem cancer şi de ce chimioterapia reprezintă cea mai bună soluţie în momentul respectiv. Transparenţa va ajuta şi la corectarea sistemelor atunci când fac greşeli.

Desigur, sunt şi domenii în care transparenţa ar putea fi un lux. Poate că nu vom insista ca un software de control al unei staţii nucleare să explice de ce opreşte reactorul. Poate că vom accepta inconvenientele provocate de pierderea temporară a energiei, dacă asta înseamnă că putem evita riscul de a avea loc o catastrofa.

Ale cui valori?

Odată ce vom începe să programăm valori cum ar fi corectitudinea în sistemele de calculatoare, ne vom confrunta apoi cu provocarea de a decide precis *ale cui* valori ar trebui să le programăm. Bineînţeles, o parte vor proveni din sistemul legal. De exemplu, o maşină autonomă va trebui să respecte orice regulă rutieră din locul în care se află. În Marea Britanie poţi să depăşeşti doar pe banda dinspre exterior şi nu trebuie să virezi niciodată când semaforul arată roşu. În Statele Unite poţi să depăşeşti pe orice bandă, şi uneori poţi să virezi când e roşu.

Dar nu toate valorile noastre sunt specificate de legi explicite. Şi chiar când sunt, unele legi trebuie încălcate. De exemplu, descoperim că maşinile autonome nu ar trebui să urmeze întotdeauna regulile de circulaţie. Un studiu din anul 2015, realizat de Institutul de Cercetare în domeniul Transporturilor din cadrul Universităţii din Michigan, a descoperit că maşinile cu pilot automat au 9,1 accidente la fiecare un milion de kilometri parcurşi, în comparaţie cu 4,1 pentru maşinile conduse de oameni[62]. Deşi au fost implicate în mai multe

[62] Cu doar unsprezece accidente ale maşinilor autonome în setul de date analizat, marja de eroare a ratei accidentelor este suficient de mare încât maşinile autonome tot ar putea, din punct de vedere statistic, să fie mai sigure decât cele

accidente, rar s-a întâmplat ca maşinile autonome să fie de vină. Ceea ce a provocat probabil multe dintre aceste accidente în mod indirect este faptul că urmau regulile prea precis. Multe dintre accidente s-au petrecut când maşinile autonome au fost lovite din spate. Dacă te loveşte cineva din spate, tehnic el este de vină aproape în toate cazurile. Mulţi şoferi umani comit mici încălcări ale legii, cum ar fi faptul că nu opresc atunci când semaforul arată galben sau trec de viteza limită atunci când depăşesc. Unele dintre aceste încălcări minore reduc numărul de accidente. Maşinile autonome vor trebui să ţină cont de, şi probabil să facă şi ele, acelaşi tip de acţiuni.

Va trebui să programăm şi valori care vor avea foarte puţin de-a face cu legile reale. De exemplu, regulile de circulaţie nu spun prea multe despre curtoazia pe care o arătăm altor şoferi, sau de obiceiurile de şofat pe care le-am dezvoltat. Dacă faci semn cu farurile înseamnă că celălalt şofer trebuie să treacă în intersecţie? Sau că e un pericol în faţă? Sau că ai o problemă cu maşina? Maşinile autonome vor trebui să înţeleagă, atunci când

---

conduse de oameni. Cu toate acestea, ne-am dori să vedem o medie mai bună pentru maşinile autonome, întrucât toţi kilometrii parcurşi de acestea au fost parcurşi pe vreme bună. Pentru mai multe detalii, vezi Brandon Schoettle & Michael Sivak, *A Preliminary Analysis of Real-World Crashes Involving Self-Driving Vehicles*, Universitatea Michigan, Institutul de Cercetare în Domeniul Transporturilor, Raportul Tehnic UMTRI-2015-34, 2015.

situaţia o cere, că pot acţiona şi ele în aceste feluri. Uneori, aceste acţiuni vor include alegeri dificile din perspectivă etică.

Problema trenului

Dilemele etice ce ţin de maşinile autonome au intrat în atenţia publicului odată cu „problema trenului”. Aceasta implică un tren scăpat de sub control. Îţi cere să iei o decizie dificilă, ce ţine de viaţă şi de moarte. Iată care este scenariul clasic.

Cinci oameni sunt legaţi de şinele de cale ferată spre care se îndreaptă trenul. Tu stai alături de oameni şi de tren, lângă maneta de schimbat macazul. Dacă tragi de această manetă, trenul o va lua pe altă linie. Sună simplu, doar că pe cealaltă linie se află o persoană. Ai două opţiuni: ori nu faci nimic şi trenul omoară cei cinci oameni de pe drumul principal, ori tragi de manetă şi faci ca trenul să schimbe direcţia, omorând o persoană. Ce alegi?

Sunt multe variaţiuni pe marginea acestui experiment mental, care implică oameni aruncaţi în faţa trenului, organe luate de la o persoană pentru a salva mai multe vieţi şi oameni închişi în loc să fie ucişi. Aceste variaţiuni pun în evidenţă distincţii etice importante, cum ar fi cea dintre acţiune şi inacţiune,

dintre efecte anticipate şi efecte intenţionate, dintre efecte directe şi posibile efecte secundare[63]. Poţi să intri pe internet şi să rezolvi şi tu astfel de probleme. Programul de la MIT Moral Machine îţi permite să rezolvi probleme de tipul celei cu trenul[64]. Ba chiar îţi permite să creezi şi tu dileme etice pentru a fi rezolvate de alţii. Scopul platformei Moral Machine este acela de crea un portret al opiniei umane cu privire la felul în care maşinile ar trebui să ia decizii când au de-a face cu dileme morale de genul acesta, folosind datele oferite de utilizatorii jocului.

Moral Machine este un bun exemplu pentru cât de bine se pricepe MIT Media Lab să atragă publicitate. Dar este foarte neclar dacă, aşa cum au sugerat unii cercetători IA, ar trebui să înzestrăm maşinile cu astfel de valori morale[65]. Chiar dacă Moral Machine a colectat

---

[63] Foarte rar se menţionează că o formă modernă a problemei cu trenul a fost introdusă cu mai bine de cincizeci de ani în urmă pentru a discuta o dilemă etică ce încă tulbură multe societăţi: moralitatea avortului atunci când viaţa unei femei este în pericol. Vezi Philippa Foot, *The Problem of Abortion and the Doctrine of the Double Effect în Virtues and Vices*, Oxford, Basil Blackwell, 1978 (a apărut iniţial în *The Oxford Review*, nr. 5, 1967).

[64] Vezi *moralmachine.mit.com*

[65] „Susţinem că luarea de decizii poate, de fapt, să fie automatizată, chiar şi în absenţa unor astfel de principii legate de adevăruri fundamentale, prin agregarea opiniilor oamenilor despre dilemele etice." Vezi Ritesh Noothigattu, Neil Gaikwad, Edmund Awad, Sohad D'Souza, Iyad Rahwan, Pradeep Ravikumar & Ariei Procaccia, „A Voting-based System for Ethical Decision Making", *Proceedings of 32nd AAAI Conference on Artificial Intelligence,* 2018.

opiniile a peste un milion de „călăi”, sunt mai multe motive pentru care nu trebuie ca maşinile să pună în practică ce au spus aceşti oameni. A răspunde la întrebări pe un website nu se compară cu momentul în care apuci de volan şi dai peste cineva în mod conştient. Vor exista oameni care vor răspunde intenţionat într-un mod pervers; am intrat odată pe Moral Machine şi am încercat să văd ce se întâmplă dacă deciziile tale omoară oameni în plus fără să fie nevoie. Moral Machine nu colectează informaţii demografice, aşa că populaţia pe care s-a făcut testarea nu este reprezentativă pentru o populaţie la scară largă. Şi chiar dacă presupunem că ar fi reprezentativă, sau ar putea să fie făcut în aşa fel încât să fie reprezentativă, chiar ne dorim ca maşinile să reflecte o medie vagă a societăţii per total?

Sunt multe argumente în favoarea ideii că până în anul 2062 va trebui să pretindem de la roboţi standarde etice *înalte* decât cele umane. Pentru că putem. Pentru că maşinile pot fi mai precise decât oamenii. Pentru că maşinile pot gândi mai rapid decât noi. Pentru că ele nu au defectele noastre.

Cu siguranţă, nu ne dorim să înzestrăm maşinăriile tocmai cu prejudecăţile pe care am vrea să le eliminăm din societate. Exact pe acestea le măsoară Moral Machine. Şi un ultim argument: ar trebui să facem ca

IA să aibă standarde etice mai presus decât oamenii pentru că, fără îndoială, pot şi trebuie să se sacrifice pentru noi.

Morală corporatistă

Probabil corporaţiile vor fi unii dintre actorii responsabili pentru a oferi valori morale maşinilor. Asta va fi o problemă, întrucât până acum s-a dovedit că multor corporaţii, mai ales companiilor din domeniul tehnologiei, le lipseşte moralitatea. Dacă nu facem ceva în privinţa asta, va fi o problemă majoră în anul 2062.

Multe dintre aceste companii de tehnologie au sediul în California, care cândva era faimoasă pentru că era locul în care să te pui pe aceeaşi frecvenţă cu ceilalţi oameni şi să ieşi din cursa de şoareci capitalistă. Dar cele mai multe dintre ele sunt finanţate de alte corporaţii de modă veche şi de bani militari. Până nu de mult mottoul Google era „Nu fi rău"; ar fi trebuit să fie un semnal de alarmă că tentaţia de a face rău exista. Nu spune ceva de genul „Fă lumea un loc mai bun" sau „Fă oamenii fericiţi"[66].

---

[66] Ca urmare a restructurării Google ca parte din conglomeratul Alphabet Inc., în anul 2015, sloganul companiei a fost schimbat de la „Nu fi rău" („*Don't be evil*") la „Fă ceea ce trebuie" („*Do the right thing*"). Noua versiune sugerează o abordare

Bineînţeles, aceste companii mari de tehnologie ne-au îmbogăţit foarte mult vieţile. Dar începem să descoperim că există nişte costuri ascunse pentru toate aceste servicii oferite gratuit. După cum se spune, când produsul nu costă nimic, adesea *tu* eşti produsul. Acestea sunt companii cu cele mai mari marje de profit din lume. Nu îşi oferă serviciile fără să primească o grămadă în schimb.

Unul dintre filosofii cei mai apreciaţi în Silicon Valley este Ayn Rand. Se pare că amalgamul ei periculos de idei libertariene şi capitaliste a contaminat multe persoane din sectorul tehnologiei. Perturbările de orice fel sunt văzute ca un lucru bun. Guvernul, per total, este văzut ca fiind un lucru rău. Putem, şi trebuie, să lăsăm piaţa să decidă. Dar piaţa nu este nici binevoitoare, nici nu are viziune pe termen lung. Sunt multe domenii în care e nevoie de reglementări pentru a ne asigura că avem cu toţii de beneficiat şi că obţinem un fel de bine comun.

Aş putea să ofer multe exemple. Ar fi foarte uşor să ne luăm de o companie precum Uber, de exemplu. Aşa

---

pozitivă a eticii corporatiste. Dar din câte ştiu eu, niciun jurnalist nu a pus întrebarea evidentă: Alphabet a făcut ce trebuie şi i-a oferit o compensaţie adecvată lui Spike Lee pentru că a folosit această expresie? Filmul său legendar din anul 1989, *Do the Right Thing*, este adesea inclus pe lista celor mai bune filme din toate timpurile. Cu siguranţă este pe lista mea. Se ştie faptul că Barack Obama şi Michelle Robinson l-au văzut la prima lor întâlnire.

că să alegem o altă ţintă. Să ne gândim la Facebook, o companie de tehnologie ce investeşte foarte mult în IA. În anul 2014, a avut loc o furtună de proteste când s-a aflat faptul că Facebook manipula în secret ca oamenii să se simtă fericiţi sau trişti. Facebook a făcut un experiment în care a ascuns postările pozitive sau negative de pe pagina principală a 689.000 de utilizatori, pentru a vedea dacă îi poate face mai fericiţi sau mai trişti. Compania nu a cerut nicio aprobare din partea vreunui comitet de etică independent pentru desfăşurarea acestui experiment, în ciuda faptului că au fost implicaţi şi doi cercetători de la Universitatea Cornell.

Eu, dacă vreau să fac un experiment pe membri ai publicului, trebuie să obţin aprobare de la comisia de etică de la universitatea mea. Trebuie să obţin acordul participanţilor care au fost informaţi cu privire la acest studiu. Trebuie să demonstrez faptul că riscurile sunt minore şi să fac în aşa fel încât să reduc orice daune posibile. Experimentul celor de la Facebook nu a făcut nimic de genul acesta. Facebook a presupus că utilizatorii şi-au dat consimţământul atunci când au spus că sunt de acord cu termenii şi condiţiile când şi-au făcut cont. Aceşti termeni şi aceste condiţii sunt, bineînţeles, atât de generale, încât nicio comisie de etică nu le-ar aproba. Şi nici înainte, nici după experiment nu

au spus nimic bieţilor utilizatori, aşa cum ar cere orice comisie de etică.

Poate că asta ar fi în regulă atunci când desfăşori un experiment de genul A sau B, când vrei să decizi ce nuanţă de albastru să alegi pentru pagina ta de web[67]. Dar nu este acceptabil atunci când încerci în mod deliberat să îi faci pe oameni mai trişti. Şi nu este acceptabil, după părerea mea, să permiţi un standard etic mai lax pentru companii decât pentru universităţi. De fapt, poate că ar trebui ca în general companiile să aibă nişte standarde mai ridicate decât universităţile, întrucât ele sunt înainte de toate în căutarea profitului, şi nu a cunoştinţelor[68].

Lumea începe să se îngrijoreze din ce în ce mai mult în ceea ce priveşte şi alte aspecte ale comportamentului companiei Facebook. Ca să mai ofer un exemplu, Facebook a introdus de curând Messenger Kids. Aceasta este o aplicaţie creată pentru copiii cu vârste între şase şi doisprezece ani. Motivul companiei este

---

[67] Testarea A sau B este un experiment statistic controlat în care compari două variante, A sau B, ca să decizi care este mai eficientă. Se ştie că Marisa Mayes, pe când era director al departamentului de inovare şi dezvoltare de produse la Google, a folosit testarea de tip A sau B pentru a alege din cele 40 de nuanţe de albastru pe cea mai potrivită culoare pentru hyperlinkuri.

[68] Având în vedere că în prezent universităţile sunt motivate mai degrabă de profituri decât de obţinerea cunoştinţelor, este oarecum ironic să spui că atunci când vine vorba de experimente pe public, companiile ar trebui să fie tratate într-adevăr ca universităţile.

clar: trebuie să găsească următoarele miliarde de utilizatori. Transformarea copiilor în utilizatori de la o vârstă fragedă este modalitatea perfectă de a face asta.

Dar există din ce în ce mai multe dovezi care indică faptul că reţelele de socializare îi pot face pe oameni să fie depresivi, anxioşi şi nefericiţi. Până şi Facebook a recunoscut că reţelele de socializare pot dăuna sănătăţii mintale. Aşa că este bine că Facebook îi încurajează pe tineri, de la o vârstă foarte impresionabilă, să folosească *mai mult* reţelele de socializare?

În Statele Unite a fost dată *Legea privind protecţia online a vieţii private a copiilor* (COPPA) în anul 1998, pentru a-i proteja pe copiii sub treisprezece ani. Această lege a împiedicat companiile de reţele de socializare să permită copiilor să-şi facă conturi pe aceste reţele, întrucât ar fi avut nevoie de acord parental pentru a face publică orice informaţie despre un copil. Facebook a susţinut că noua sa aplicaţie Messenger Kids a fost aprobată de un comitet consultativ de experţi, ce proveneau din domenii cum ar fi dezvoltarea copiilor, siguranţa în media şi siguranţa online. Dar revista *Wired* a descoperit mai târziu că majoritatea acestor experţi fuseseră anterior finanţaţi de Facebook.

Nu ar trebui atunci ca Facebook să depună mai mult efort să îi *scoată* pe cei care au sub treisprezece ani de pe

platforma lor, în loc să facă astfel încât să le fie mai uşor să se alăture? Mark Zuckerberg a promis la un moment dat chiar că va lupta ca COPPA să fie înlăturată[69]. În anii ce vor urma, vom privi în urmă cu nostalgie, gândindu-ne la vremurile acestea mai inocente, când copiii noştri încă nu erau expuşi la caruselul reţelelor de socializare?

Facebook a declarat că misiunea lor este următoarea: „Dă-le oamenilor puterea de a construi o comunitate şi fă ca lumea să fie mai unită". În anul 2017, Propublica a descoperit că algoritmii companiei vindeau reclame care se adresau persoanelor antisemite. Facebook a vândut anunţuri pentru locuri de muncă ce discriminează împotriva oamenilor în vârstă şi face reclamă la propriile joburi discriminând în favoarea tinerilor. Este greu să înţelegi cum astfel de activităţi de afaceri construiesc comunitatea şi fac ca lumea să fie mai unită.

Eliminarea părtinirii

Ar putea suna ideal să eliminăm orice favoritism, dar nu este posibil. Dacă Amazon îţi recomandă o carte, *match. Com* îţi propune o întâlnire, sau *monster.com* îţi

[69] Michael Lev-Ram, „Zuckerberg: Kids Under 13 Should Be Allowed on Facebook", *Fortune*, 20 mai 2017.

sugerează o nouă slujbă care ţi-ar putea plăcea, asta constituie o părtinire care creează o preferinţă pentru anumită carte, întâlnire sau job. Ba chiar o bună parte din învăţarea mecanică constă în a decide care este cel mai dezirabil tip de părtinire cu care să fie înzestrat un program. Această părtinire, numită adesea părtinire inductivă, este setul de presupuneri folosite pentru a prezice outputuri, atunci când apar date de intrare care nu au mai fost întâlnite anterior.

Favoritismul poate fi chiar dezirabil. Poate ne dorim ca procesul de admitere la facultăţi să fie în favoarea celor din cartierele sărace. Sau ca deciziile de împrumut să fie în favoarea celor care cel mai probabil vor plăti. Sau ca traducerea mecanică să pornească cu o preferinţă împotriva sexismului înglobat în corpusul ei de limbi. Sau ca maşinile autonome să îi lase pe pietoni şi pe biciclişti să treacă.

Şi chiar există instrumente care pot muta părtinirea dintr-o direcţie nedorită într-una dezirabilă. Putem, de exemplu, să încercăm să îmbunătăţim acurateţea algoritmilor. Poate că trebuie să îi antrenăm cu mai multe date, sau să le adăugăm caracteristici noi, sau să schimbăm modelul pentru a le îmbunătăţi acurateţea. O altă opţiune este să punem pe lista neagră anumite răspunsuri. Aşa cum am văzut mai devreme, Google a pus pe lista neagră cuvântul „gorilă” şi acesta nu mai

este o etichetă disponibilă în aplicaţia sa de poze. Problema cu aceste liste negre este că, foarte des, este dificil să te asiguri că sunt complete. Bineînţeles, poţi să întorci soluţia pe dos şi să faci o listă albă de răspunsuri acceptabile. Dar atunci s-ar putea să ratezi o serie de răspunsuri noi şi utile.

O altă cale prin care putem combate favoritismul este prin ştergerea trăsăturilor din setul de date care introduc un fel de preferinţă nedorită. Dacă nu vrei ca deciziile pentru împrumuturi să depindă de rasă, atunci nu include rasa ca dată de intrare. Dar eliminarea rasei s-ar putea să nu fie de ajuns; aşa cum am văzut, ar putea fi alte aspecte din cadrul setului de date, cum ar fi codul poştal, care sunt în corelaţie strânsă cu rasa. Putem elimina aceste trăsături corelate, dar dacă eliminăm prea multe, s-ar putea să afectăm acurateţea. De asemenea, putem să schimbăm setul de date cu totul. Dacă setul pe care s-a făcut învăţarea cuprindea prea mulţi bărbaţi, am putea alege să creştem numărul de femei din baza de date. Poate fi aceasta modificată astfel încât să aibă aceleaşi date demografice ca populaţia generală?

În cele din urmă, ultimul instrument pe care-l putem folosi în cazul părtinirii este conştientizarea. Nu există nicio cale perfectă prin care să detectăm şi să schimbăm preferinţele din sistemele IA. Dar fără să conştientizăm

faptul că acea preferinţă ar putea fi prezentă, nu există nicio posibilitate de schimbare.

Epoca de aur a filosofiei

Când a avut loc epoca de aur a filosofiei? Când au fost puse bazele filosofiei, în timpul lui Socrate, Aristotel sau Platon? Sau în timpul lui Descartes, care este considerat adesea ca fiind părintele filosofiei moderne vestice? Sau poate că în timpul lui Confucius şi al discipolilor săi, ale căror idei ne influenţează şi în ziua de azi? De exemplu, ideea de a nu le face altora ce nu vrei să ţi se facă ţie. Eu presupun, totuşi, că epoca de aur a filosofiei acum este pe cale de a începe.

Următoarele decenii ar putea constitui o perioadă prosperă pentru filosofie, întrucât vom trata cu seriozitate alegerile problematice din punct de vedere etic. Având în vedere cât de literale sunt calculatoarele, va trebui să fim mai precişi ca niciodată în privinţa valorilor noastre, în timp ce oferim sistemelor IA abilitatea de a lua decizii care vor avea un impact asupra noastră. Până în anul 2062, fiecare companie mare va avea nevoie de un nou fel de CFO – Coordonator Filosofic Oficial – care va ajuta compania să decidă cum vor acţiona sistemele sale IA. Iar domeniul *eticii computaţionale* va înflori, pe măsură ce

ne vom gândi cum să construim sisteme care să urmeze valorile stabilite de comun acord.

De curând, un prieten m-a întrebat cum să-şi convingă copilul să nu opteze pentru filosofie la facultate, ci pentru ceva „mai practic”. Răspunsul meu a fost să îl felicit pe copil pentru că a ales acest domeniu. Avem nevoie disperată de filosofi în afaceri, în guvern şi oriunde altundeva. Fără ei, va fi imposibil să ne asigurăm că sistemele IA ale anului 2062 vor oglindi valorile umane. Şi să ne asigurăm că în cele din urmă Homo digitalis va fi mai etic decât Homo sapiens.

0007
Sfârşitul egalităţii

Egalitatea este una dintre valorile umane aflate în pericol din cauza schimbărilor tehnologice. Bineînţeles, egalitatea nu va dispărea. Încă de la începuturile societăţii a existat inegalitate. Nicio societate nu a fost cu adevărat egală. Întotdeauna au fost oameni care s-au născut mai bogaţi şi care au avut oportunităţi mai bune. Dar perioada scurtă în care inegalitatea s-a aflat în scădere, care a precedat cele două războaie mondiale, s-a terminat, şi inegalitatea este din nou în creştere. Până în anul 2062, am putea vedea inegalităţi majore în cadrul societăţii noastre. Aşa că probabil

titlul acestui capitol ar trebui să fie acesta, mai puţin atrăgător: „Sfârşitul scăderii inegalităţii”.

Economişti precum Thomas Piketty au demonstrat convingător faptul că inegalitatea este în creştere în economiile capitaliste atunci când rentabilitatea capitalului depăşeşte rata de creştere economică. În acel caz, cei care deţin capital au de câştigat, spre deosebire de cei care deţin doar propria lor muncă. Inegalitatea s-a aflat în creştere în cea mai mare parte a istoriei noastre economice. Iar asta nu doar în sistemul capitalist. Nici comunismul nu s-a descurcat mai bine în practică, deoarece a favorizat câteva persoane şi a eşuat să-i ridice pe cei mai săraci[70].

Alte tendinţe, cum ar fi globalizarea şi criza financiară care nu se mai termină niciodată, contribuie şi ele la această creştere a inegalităţii. Din nefericire, IA va duce la o creştere a inegalităţii şi mai mare, averea şi puterea ajungând în mâinile elitei tehnologice – asta dacă nu facem ceva pentru a corecta acest lucru în viitorul apropiat.

---

[70] Bineînţeles, există şi alte opţiuni în afară de capitalism şi comunism. Economia de piaţă socialistă a Chinei este un caz interesant, la care mă voi referi mai în detaliu ceva mai încolo. China a redus sărăcia mai mult decât oricare altă economie; în ciuda acestui lucru, inegalitatea din China a crescut şi ea semnificativ.

Viaţa nu a fost niciodată mai bună

Cu o sută de ani în urmă speranţa de viaţă din Statele Unite, Australia şi Marea Britanie era în jurul vârstei de 55 de ani. În prezent, este de peste 80 de ani. A existat o perioadă în care speranţa mea de viaţă ca bărbat adult creştea în fiecare an cu un an în plus. Era un ritm plin de viaţă pentru mine.

Sărăcia extremă a scăzut sub 10% din populaţia globului pentru prima dată. În anii 1900 sărăcia afecta peste 80% din oamenii de pe planetă. Educaţia a fost una dintre cauzele majore care au dus la această schimbare. Cu câteva sute de ani în urmă, în jur de 15% din populaţie ştia să citească. Azi, 80% din oameni ştiu să citească şi să scrie. Chiar mai încurajator, 90% din populaţia lumii cu vârste sub 25 de ani ştie să citească. Această reducere a sărăciei extreme a avut un impact imens. Acum e mai probabil să mori de obezitate decât de subnutriţie.

Cei mai mulţi dintre noi nu ne dăm seama de acest lucru, dar este şi perioada cu nivelul cel mai scăzut de violenţă din întreaga noastră istorie. Rata omuciderilor anuale din Londra a scăzut de la peste 50 de oameni la fiecare 100.000, cât era în secolul al XV-lea, la mai puţin de 2 oameni la fiecare 100.000. Şi în ciuda genocidelor teribile din Bosnia, Rwanda, Siria şi din alte locuri, rata

mortalităţii din cauza războaielor civile a scăzut de 10 ori în ultimii 50 de ani.

Dar în timp ce calitatea vieţii a crescut peste tot în lume pentru cei care aveau cea mai proastă situaţie materială, distanţa faţă de cei care au fost suficient de norocoşi încât să câştige la loteria vieţii se măreşte rapid. Cei mai bogaţi 500 de oameni de pe planetă şi-au crescut averea cu peste un trilion de dolari doar în anul 2017. Şi cei mai bogaţi 8 miliardari ai lumii controlează acum aceeaşi avere ca jumătatea săracă a populaţiei de pe glob. Viaţa nu a fost niciodată mai bună ca acum – în special pentru oamenii bogaţi.

Cel mai rău loc în care poţi trăi

Probabil că ţara în care inegalitatea este cea mai evidentă este SUA. Niciuna dintre cele 35 de naţiuni membre ale Organizaţiei pentru Cooperare şi Dezvoltare Economică nu este la fel de inegală ca Statele Unite, şi niciuna nu a experimentat o creştere atât de bruscă în inegalitate. În SUA, cei mai bogaţi oameni ce reprezintă 1% din populaţie aproape şi-au dublat venitul din anul 1980, de la 11% la 20% din total.

Ca să punem lucrurile în perspectivă, ponderea veniturilor în Danemarca în cazul primilor 1% a crescut de la 5% la doar 6% în această perioadă. În Olanda nu a

existat nicio creştere de la nivelul de 6%. Presupun că nu este o coincidenţă faptul că Danemarca şi Olanda sunt adesea în topul clasamentului celor mai bune ţări în care să trăieşti. Sunt şi alte ţări care au avut creşteri semnificative în cazul veniturilor celor mai bogaţi 1% cetăţeni. De exemplu, Anglia a avut parte de o creştere a veniturilor primilor 1% de la 6% la 14%, iar Canada de la 9% la 14%.

Dar nici o altă ţară nu a mai avut parte de o creştere atât de mare ca a Statelor Unite ale Americii. Şi niciuna nu a pornit de la o bază atât de inegală.

Şi nu doar cei săraci sunt lăsaţi în urmă. Oamenii din clasa de mijloc nu au nici ei parte de prosperitatea în creştere. Potrivit Institutului de Politică Economică, venitul median pe oră din Statele Unite s-a schimbat foarte puţin de-a lungul deceniilor. În anul 2016, veniturile mediane au crescut de la 16,74 de dolari pe oră în anul 1973 la doar 17,86 de dolari pe oră. Având în vedere că îngrijirile medicale costă din ce în ce mai mult şi siguranţa unui loc de muncă este în scădere, mulţi dintre cei din clasa mijlocie se simt strânşi cu uşa, şi pe bună dreptate.

Economia „picurării”

Un argument care este folosit adesea pentru a ne

încuraja să acceptăm faptul că cei bogaţi devin din ce în ce mai bogaţi este că averea lor „va picura” şi va îmbunătăţi vieţile tuturor. Mai este şi argumentul care spune că taxarea celor bogaţi va încetini dezvoltarea şi inovaţia. Sunt foarte puţine dovezi care să susţină cele două argumente. Într-adevăr, sunt multe aspecte care demonstrează opusul.

O analiză a Fondului Monetar Internaţional sugerează că creşterea cotei veniturilor care revin celor săraci şi clasei mijlocii accelerează dezvoltarea, în timp ce creşterea cotei veniturilor primilor 20% de fapt încetineşte dezvoltarea. Când cei bogaţi se îmbogăţesc şi mai mult, beneficiile nu se scurg şi spre cei săraci. Când cei săraci se îmbogăţesc, la fel fac şi cei bogaţi[71].

Un experiment interesant relevant pentru economia „picurării” a avut loc în 2012. Guvernatorul Kansasului, Sam Brownback, a implementat un program de reducere generoasă a taxelor pentru oamenii de afaceri şi pentru cei avuţi, şi de reducere a taxelor mai puţin generoasă pentru cei care câştigă mai puţin. Cinci ani mai târziu, economia statului era într-o stare teribilă. Se pierdeau mii de locuri de muncă anual. Statul tăia din fondurile de pensii şi reducea bugetele

---

[71] Vezi Era Dabla-Norris, Evridiki Tsounta, Kalpana Kochhar, Frantisek Ricka & Nujin Suphaphiphat, *Causes and Consequences of Income Inequality: A Global Perspective*, raport tehnic, Fondul Monetar Internaţional, iunie 2015, SDN/15/13.

pentru universităţi, pentru MedicAid şi pentru alte servicii. În anul 2017 Kansas s-a dat bătut şi a inversat reducerile de taxe.

În aceeaşi perioadă, California a încercat să facă opusul. În luna noiembrie a anului 2012, alegătorii din acest stat au fost de acord cu Proposition 30, care a ridicat temporar taxele pe venituri impuse de stat pentru cei mai bogaţi rezidenţi din California şi taxele pe vânzări. Profiturile obţinute au fost folosite pentru a finanţa şcoli şi pentru a plăti datorii de 27 de milioane de dolari. De atunci, California s-a bucurat de una dintre cele mai puternice creşteri economice din toate statele americane. Desigur, alţi factori – cum ar fi sectorul IT – au contribuit şi ei la creşterea economică a Californiei. Dar se pare că taxarea celor bogaţi nu a avut vreun dezavantaj.

Înainte a fost altfel

A fost o perioadă ce a urmat după al Doilea Război Mondial în care lucrurile stăteau altfel. Inegalitatea a scăzut şi mobilitatea socială a crescut. Introducerea ajutoarelor sociale, a dreptului muncii şi a sindicatelor, a educaţiei universale, precum şi măsuri locale, cum ar fi *Actul Veteranilor* din Statele Unite şi *Actul Serviciilor Naţionale de Sănătate* din Regatul Unit, au creat

condiţiile necesare unei schimbări majore şi unei perioade care acum pare să fi avut parte de o reducere a inegalităţii destul de neobişnuită.

Aceste schimbări au fost provocate de câteva şocuri mari pentru societate: două războaie mondiale, Marea Depresiune interbelică şi ameninţările iminente ale comunismului, Războiului Rece şi spectrului anihilării nucleare. Poate că provocări precum criza financiară globală şi încălzirea globală vor reprezenta şocurile necesare pentru a reforma societatea pentru revoluţia ce urmează să fie adusă de IA?

N-aş prea crede. Politicienii nu au nici curajul, nici viziunea pentru a acţiona suficient de hotărât. Şi sistemele politice nu îi încurajează să fie curajoşi şi vizionari. Va fi nevoie de mult mai mult decât tipărirea de bani pentru a se crea condiţiile prielnice unui rezultat pozitiv. Trebuie să luăm în considerare schimbări radicale la nivelul ajutoarelor sociale, al sistemului de taxare, al sistemului educaţional, al dreptului muncii şi chiar al instituţiilor politice. Nu cred că tratăm aceste discuţii cu destulă seriozitate.

Probabil este prea târziu să prevenim schimbările climatice majore. Acum trebuie să încercăm să ne descurcăm cu efectele lor. În mod similar, mă tem că nu vom acţiona suficient de rapid astfel încât să prevenim ca perturbările produse de tehnologie să afecteze

societatea. Scriind această carte, unul dintre obiectivele mele este să trag un semnal de alarmă şi să îndemn lumea să facă nişte schimbări rapid.

Inegalitate corporatistă

Bogăţia nu este concentrată numai în mâinile celor bogaţi. Este şi în soldurile bancare ale câtorva corporaţii extrem de puternice. Din nou, dacă nu vom face ceva în privinţa asta, până în anul 2062 rezultatul va fi unul sumbru.

Pieţele digitale sunt deseori monopoluri naturale. Câştigătorul ia totul. Ne dorim şi avem nevoie doar de un singur motor de căutare, o singură aplicaţie de mesaje, un singur serviciu de reţea de socializare. Şi cu excepţia Chinei, unde competiţia a fost impusă, avem într-adevăr un singur motor de căutare dominant, o aplicaţie de mesaje dominantă şi o reţea de socializare dominantă.

În ultimele trei luni ale anului 2007, companiile listate public cu cel mai mare capital de piaţă au fost Petrochina, Exxonmobil, General Electric şi China Mobile. Zece ani mai târziu, primele patru companii sunt toate de tehnologie: Apple, Alphabet (compania din care face parte Google), Microsoft şi Amazon. În 2007 doar Microsoft se afla între primele zece.

Companiile de tehnologie precum Google au considerat cândva că trebuie să rămână competitive. O companie nouă mai bună ar putea apărea în orice moment, iar utilizatorii şi-ar putea schimba motorul de căutare preferat cât ai zice peşte. Dar nu mai este cazul de aşa ceva. Nici o altă companie nu mai are datele sau resursele financiare necesare pentru a concura cu Google. O corporaţie precum Google îşi permite să plătească miliarde de dolari ca să cumpere orice companie nou apărută care îi ameninţă poziţia. Cu câţiva ani în urmă, Google cumpăra câte o companie săptămânal. Şi dacă proprietarii refuzau să o vândă, Google construia un serviciu similar şi îl oferea gratuit, astfel eliminând orice competiţie. David nu are nicio şansă în faţa unui astfel de Goliat.

Conflictele cu Goliaţi

Guvernele au introdus adesea reglementări pentru a asigura că există competiţie. În timp ce legi care impuneau preţuri corecte la cereale sunt atestate încă de pe vremea romanilor, eficientizarea datorată producţiei la scară largă adusă de revoluţia industrială a creat o nevoie şi mai mare de intervenţie. Prima lovitură din acest război continuu a fost spargerea monopolului petrolier.

În anii 1900, Standard Oil controla peste 90% din petrolul rafinat din Statele Unite. John D. Rockefeller a fost fondatorul, preşedintele şi deţinătorul majoritar de acţiuni al companiei. Este considerat a fi cel mai bogat om din istoria modernă. În anul 1911, în urma protestelor populaţiei, Curtea Supremă a împărţit Standard Oil în 34 de companii mai mici, prin *Sherman Antitrust Act* din anul 1890.

Au urmat şi alte atacuri asupra Goliaţilor. În 1911, Curtea Supremă a Statelor Unite a atacat monopolul din domeniul tutunului, împărţind American Tobacco Company în patru firme mai mici. În anul 1892, curtea s-a ocupat de monopolul din telecomunicaţii, şi a împărţit AT&T în şapte companii regionale operate de Bell, plus compania-mamă, acum mult redusă. Mai recent, Curtea Supremă a pus ochii pe monopolurile din domeniul tehnologiei, şi anume pe Microsoft, dar nu a avut succes. Dincolo de oceanul Atlantic, Comisia Europeană a avut mai mult succes ocupându-se de comportamentul anticoncurenţial din sectorul IT. În anul 2017, Uniunea Europeană a amendat Google cu o sumă-record de 2,7 miliarde de dolari pentru concurenţă neloială.

În Statele Unite, legislaţia în sprijinul concurenţei s-a concentrat pe preţurile plătite de consumatori. Cu toate acestea, comportamentul anticompetitiv nu are doar un

impact direct asupra preţurilor. Adesea, preţurile de pe piaţă nu includ externalităţi cum ar fi costul poluării, în special pe termen scurt. Există de asemenea o presiune în creştere pentru a fi reglementate şi monopolurile de date. Din păcate, nici măcar amenzile-record nu par să aibă efectul scontat asupra acţiunilor marilor companii de tehnologie.

Alte măsuri au dat şi ele greş. În 2011, Google a plătit 700 de milioane de dolari pentru ITA, o companie care oferă software pentru rezervările de zboruri făcute de companiile aeriene. Departamentul de Justiţie a aprobat vânzarea, cu condiţia ca Google să păstreze software-ul accesibil pentru alte afaceri timp de cel puţin cinci ani. Google a aşteptat doar doi ani înainte să anunţe că va închide accesul la software în 2018. Ten Benson, un fost angajat la Google, a scris pe Twitter: „A fost executat un ecosistem întreg de companii aeriene noi printr-o singură tăietură cu pixul".

Lăcomie corporatistă

Companiile de tehnologie lucrează de multe ori pe pieţe care aduc profituri extraordinare. Nu costă aproape nimic ca bunurile digitale să fie reproduse, iar serviciile digitale pot să fie extinse rapid şi ieftin. Alphabet, spre exemplu, obţine profit net de peste 20%

din venitul său anual de 110 miliarde de dolari. Prin comparaţie, Walmart obţine profit de mai puţin de 3% din venitul său de 485 de miliarde de dolari.

O parte din averea acestor corporaţii de IT mari apare în urma refuzului lor de a plăti taxe. Un raport publicat de Comisia Uniunii Europene în septembrie 2017 a descoperit că o afacere digitală din UE care operează la nivel internaţional plăteşte o taxă efectivă de 10%, în comparaţie cu o rată de 23% plătită de companiile internaţionale. Multe companii de IT plătesc chiar şi mai puţin.

În anul 2016, Amazon a plătit taxe de doar 15 milioane de lire pentru venituri europene de 19, 5 miliarde de lire. Asta nu înseamnă nici măcar 0,1% din venituri. Veniturile Facebook din Anglia din anul 2016 au crescut la 842 de milioane de lire, comparativ cu anul anterior, când au fost de 210,8 milioane de lire, dar cumva taxele lor au crescut doar de la 4,2 milioane de lire la 5,1 milioane. Asta înseamnă mai puţin de 1% din venituri. Este remarcabil că veniturile companiei s-au cvadruplat, dar taxele plătite de ea au crescut doar cu un sfert.

Bineînţeles, nu doar companiile tehnologice plătesc taxe mici. Evaziunea fiscală s-a transformat într-un joc jucat de multe corporaţii. De exemplu, în cele mai recente conturi ale sale din Australia, IKEA a raportat

vânzări de 1,16 miliarde de dolari australieni. Dar după ce contabilii şi-au făcut bine treaba, se zice că IKEA nu a plătit decât 289.000 de dolari australieni către fiscul australian. Lumea s-ar putea întreba de ce IKEA a depus atât de mult efort pentru a vinde mobilă în colete plate în valoare de peste un miliard de dolari în Australia, când este clar că acest tip de comerţ nu este prea profitabil.

Dar este adevărat că în general companiile de tehnologie sunt printre cele mai agresive atunci când vine vorba de evitarea taxelor. În Irlanda, Comisia Europeană a descoperit că Apple a plătit taxe la o rată de doar 0,005% către autorităţile irlandeze în anul 2014, cu mult sub rata de taxare a corporaţiilor de 12,5%. Comisia a decis că Apple trebuie să plătească 13 miliarde de euro în impozite. Drept rezultat, acum avem o situaţie bizară în care guvernul irlandez se luptă cu decizia Comisiei, încercând să *nu* primească miliarde de euro în taxe. Guvernul irlandez joacă un joc de lungă durată, sperând că Irlanda va rămâne unul dintre paradisurile fiscale din Europa.

Regatul Unit şi Australia au introdus amândouă „taxe Google" speciale pentru a încerca să forţeze companiile de tehnologie să plătească o taxă rezonabilă. Pare corect ca aceste corporaţii să contribuie cu ceva în locurile de unde provin veniturile

lor. Într-adevăr, dacă s-ar gândi mai bine, ar vedea că, la urma urmei, nu este în interesul lor să secătuiască averea clienţilor lor.

Aceste corporaţii tehnologice nu au nevoie să genereze profituri atât de mari pentru a continua să crească. Majoritatea nu plătesc dividende, aşa că nu au nevoie să fie prea profitabile ca să îşi recompenseze acţionarii în mod direct. Profiturile lor sunt folosite adesea pentru a-şi cumpăra acţiunile înapoi. Aceasta este o opţiune leneşă. Spune: „Nu putem face nimic util cu toţi aceşti bani". Şi are ca efect creşterea preţului acţiunilor, ceea ce este în interesul directorilor care deţin acţiuni. Aceste lucruri trebuie să fie reglementate.

Distrugerea companiei Uber

Cum au luat-o lucrurile la vale? Ce s-a întâmplat cu visul digital? Să luăm taxiurile ca exemplu. Cu un deceniu în urmă, în multe locuri, piaţa taxiurilor nu era competitivă. În multe oraşe nu se putea găsi un taxi într-o zi ploioasă. Licenţele de taxiuri erau adesea ridicol de scumpe. Uber se presupunea că avea să rezolve aceste probleme, dar a ajuns să înlocuiască un sistem defect cu altul.

Uber poate fi criticat din foarte multe puncte de vedere. Avem şoferii de Uber care sunt forţaţi să

trăiască în maşinile proprii pentru că sunt plătiţi cu foarte puţin. Şi se pare că Uber nu are scrupule legate de faptul că nu respectă legea sau că îşi spionează companiile rivale. Nici nu consideră că este nevoie să le spună utilizatorilor săi că datele lor au fost furate. Lăsând la o parte comportamentul lor imoral, există totuşi o problemă fundamentală cu Uber: fură majoritatea avuţiei din sistem.

Se presupunea că internetul avea să elimine verigile intermediare, permiţându-ne să construim pieţe digitale care lucrează mai eficient. De ce este atât de dificil să îi conectezi pe oamenii care au maşini disponibile cu cei care au nevoie de transport, astfel încât *toată lumea* să aibă de câştigat? Este corect că Uber extrage o valoare atât de mare din sistem? Nu ar fi mai bine, spre exemplu, dacă am stabili o cooperativă în care proprietarii maşinilor şi clienţii lor să poată împărţi valoarea generată?

Există două motive pentru care acest lucru nu se întâmplă. Unul este tehnologic (şi acum poate fi reparat). Problema tehnologică este că nu vrei să intri în maşina unei persoane oarecare, sau nu vrei ca o persoană oarecare să-ţi între în maşină. Aşadar vei avea nevoie de un sistem bazat pe reputaţie, astfel încât şoferii şi pasagerii să aibă încredere unii în alţii. Până de curând, un sistem bazat pe reputaţie însemna că

aveai nevoie de o autoritate centrală care putea să menţină istoricul comportamentului şoferilor şi al pasagerilor. Iar Uber era această autoritate centrală. Dar nu mai este cazul acum. Cu tehnologia blockchain am putea construi un sistem bazat pe reputaţie descentralizat, şi astfel nu ar mai fi nevoie de un intermediar ca Uber.

Rămâne obstacolul financiar. Uber nu are nevoie să producă bani. A strâns şi va continua să strângă sume mari de bani de pe piaţa de capital de risc. Prin urmare, piaţa de *ride-sharing*, pe care activează companii ca Uber, nu mai este una competitivă. Nu mai contează care companie oferă servicii mai bune. Câştigătorul în ultimă instanţă nu este nici consumatorul, nici şoferul de taxi. Câştigătorul este pur şi simplu fondul de investiţii care are cele mai adânci buzunare. O cooperativă nu poate concura împotriva unei afaceri care nu are nevoie să facă profit, şi cu siguranţă nu poate concura împotriva unei companii care nu are nevoie nici măcar să iasă pe zero, ci poate pierde bani fără nicio problemă. În 2016, Uber a avut pierderi 2,8 miliarde de dolari, în timp ce a obţinut venituri de 6,5 miliarde de dolari. Asta înseamnă aproape un dolar pentru fiecare 2,32 dolari intraţi. Finanţatorii companiei Uber plătesc aproximativ o treime din fiecare cursă pe care o faci.

Cercetare privată

Poate că eşti pregătit să tolerezi capitalismul pur şi dur, având în vedere toate beneficiile pe care corporaţii precum Google şi Apple le-au adus în vieţile noastre. Gândeşte-te, de exemplu, de câte ori pe zi depinzi de Google, sau de iphone-ul din buzunar. Dar o bună parte din tehnologia care ne face viaţa mai uşoară nu este creată de corporaţii. Internetul a fost finanţat de DARPA (*Deferise Advanced Research Projects Agency*, Agenţia pentru proiecte de cercetare avansată în domeniul apărării), o instituţie guvernamentală americană. World Wide Web a fost inventat de Organizaţia Europeană pentru Cercetare Nucleară (cunoscută şi drept CERN), un laborator de fizică finanţat de mai multe guverne. Şi o mare parte din tehnologia folosită în iphone a provenit din cercetări subvenţionate de guvern. Ecranele cu touchscreen, GPS-ul, internetul şi chiar şi tehnologia din spatele lui Siri, pentru toate acestea nu au plătit fondurile de investiţii din Silicon Valley, ci taxele oamenilor.

Cercetarea necesită o viziune pe termen lung. Presupune multe pariuri, dintre care cele mai multe nu vor fi câştigate. Şi adesea contribuie la binele public, şi nu la binele vreunei companii anume. Ştiinţa nu este

ceva ţinut secret. Ca oameni de ştiinţă, facem totul public pentru ca toată lumea să beneficieze. Desigur, corporaţiile contribuie la ecosistemul cercetării. Şi instituţii precum sistemul de patentare ajută oamenii de ştiinţă să publice şi să primească totuşi o binemeritată răsplată pentru munca lor.

În ultimii ani, marile companii din domeniul tehnologiei, cum ar fi Microsoft, Google şi Facebook, şi-au întărit legăturile cu universităţile. Mulţi dintre specialiştii seniori recunosc faptul că inovaţiile provin adesea din universităţi, nu din laboratoare corporatiste. Admit şi faptul că aproape toţi cercetătorii din laboratoarele de cercetare şi inovare ale companiilor de tehnologie provin din universităţi.

Este adevărat că corporaţiile, şi nu fondurile guvernului, au alimentat o mare parte din progresul enorm care a avut loc recent în domeniul învăţării profunde. Dar bazele acestui progres pot fi urmărite până la finanţările guvernamentale. Timp de mulţi ani, Institutul Canadian pentru Cercetări Avansate i-a finanţat pe Geoff Hinton de la Universitatea din Toronto şi pe Yoshua Bengio de la Universitatea din Montreal. Hinton şi Bengio s-au ocupat cu cercetările legate de reţelele neurale, un subiect ce părea demodat la vremea respectivă. Pariul canadienilor a dat roade în mod spectaculos, întrucât în prezent Canada este o

pepinieră de companii noi în domeniul învăţării profunde.

Corporaţia modernă

Este uşor să uităm că de fapt corporaţia modernă reprezintă produsul precedentei revoluţii din economie, revoluţia industrială. Sunt foarte puţine companii în prezent, majoritatea fiind bănci şi edituri, care datează de acum 300 de ani. Cele mai multe corporaţii au fost înfiinţate mult mai recent. Vârsta medie a companiilor incluse de indicele S&P 500 este de doar 20 de ani.

Până la urmă, corporaţia modernă este o instituţie creată de om, făcută în mare parte pentru a permite societăţii să profite în urma unor schimbări tehnologice. Răspunderea limitată le permite directorilor corporaţiilor să îşi asume riscuri, fără să ajungă la riscuri personale. Piaţa de acţiuni şi piaţa de obligaţiuni oferă corporaţiilor acces la fonduri care le ajută să investească în tehnologii şi pieţe noi. Iar transferabilitatea acţiunilor oferă continuitate corporaţiilor, astfel încât acestea să poată creşte pe parcurs.

Companii precum Dupont (fondată în anul 1802), General Electric (1892), Ford Motor Company (1903) şi

IBM (1911) sunt clar produse ale revoluţiei industriale. Dar în ziua de azi o problemă extrem de importantă este că, deşi multe corporaţii beneficiază foarte mult de pe urma progreselor recente din domeniul tehnologiei, o mare parte din restul societăţii nu are aceste beneficii. Multe companii tehnologice sunt structurate astfel încât să ofere tratament preferenţial fondatorilor lor, şi nu acţionarilor. Şi chiar şi în cazurile în care fondatorii nu sunt astfel privilegiaţi, structura conducerii corporaţiei le permite directorilor executivi să îşi maximizeze profiturile. Acestea nu sunt în proporţie corectă cu profiturile care revin acţionarilor sau, mai general, cu valoarea care revine societăţii.

Pe scurt, unele dintre aceste companii de tehnologie au devenit la fel de valoroase şi de puternice ca nişte ţări mici. Să luăm ca exemplu Apple, care în prezent valorează 850 de miliarde de dolari, şi care în curând s-ar putea să devină prima companie de un trilion de dolari. Să comparăm cu una dintre ţările mai bogate, dar mai mici din lume, Luxemburg. O regulă orientativă în contabilitate este că un bun valorează de zece ori mai mult decât venitul său anual. Venitul anual al Luxemburgului este produsul său intern brut, care reprezintă valoarea totală a tuturor bunurilor şi a serviciilor produse de locuitorii ţării Luxemburg de-a lungul unui an. Asta înseamnă în jur de 60 de miliarde

de dolari. Făcând aceste estimări, am putea spune că Luxemburg valorează în jur de 600 de miliarde de dolari – sau mai puţin decât un Apple.

Corporaţia secolului XXI

Având în vedere că de fapt corporaţia modernă este o invenţie umană şi, în parte, un produs al Revoluţiei Industriale, poate că este timpul să ne gândim la cum am putea reinventa ideea de corporaţie pentru a se adecva la revoluţia ce urmează să vină. În special, cum ne putem asigura că în anul 2062 corporaţiile vor contribui mai mult la binele comun?

Sunt câteva soluţii care ar putea fi utile pentru corporaţiile secolului XXI. Prima este să limitezi puterea directorilor executivi (CEO) şi să faci astfel încât angajaţii şi acţionarii să fie mai bine reprezentaţi. Germania oferă un exemplu bun cu privire la beneficiile acordării unei puteri sporite angajaţilor. Companiile din Germania cu peste 2000 de angajaţi sunt obligate să aibă un consiliu de supraveghere, jumătate din membrii acestuia fiind angajaţi. Acest consiliu de supraveghere stabileşte salariile directorilor şi angajează şi concediază CEO-ul şi alţi directori.

Există şi posibilitatea limitării raportului dintre salariul directorului executiv şi al celor mai prost plătiţi

angajaţi ai companiei. În Statele Unite, această proporţie a crescut de şase ori, în medie, în ultimele trei decenii. Un CEO lucrează de şase ori mai mult acum? Iar în timp ce salariile lor au crescut semnificativ, salariile medii abia dacă s-au modificat vreun pic în tot acest timp. Am putea să limităm şi acţiunile oferite directorilor executivi, dacă aceste stimulente nu sunt oferite şi altor angajaţi.

Al doilea ingredient al reformei corporaţiei moderne ţine de stimularea dezvoltării societăţilor mutuale şi cooperativelor. Asta ar putea include: stimulente financiare cum ar fi reducerea taxelor; împrumuturi guvernamentale mai ieftine şi acces favorizat la capital; contracte pentru achiziţii acordate preferenţial de instituţiile guvernamentale, în favoarea acestor companii; şi cerinţe mai relaxate privind lichidităţile decât cele pentru companiile listate public (a căror conducere ar putea fi tentată să îşi asume mai multe riscuri, deoarece se joacă cu banii altora).

O a treia soluţie ţine de reformele privind taxarea, care să oblige corporaţiile să plătească mai multe taxe în locurile de unde îşi obţin averea. Această situaţie va fi în interesul tuturor pe termen lung. Nu este sustenabil să storci o ţară de avere. Corporaţiile beneficiază în urma unei societăţi ordonate, în care pot opera. Drumuri, spitale, şcoli, transport şi altele

asemenea trebuie finanţate, ca pieţele respective să existe.

O a patra variantă este reforma dreptului muncii. Într-o economie a precarităţii muncii, angajaţii trebuie să aibă drepturi mai multe, astfel încât să se bucure de protecţia pe care corporaţiile o ofereau în trecut. Angajaţii trebuie să aibă dreptul să se poată îngriji când sunt bolnavi, să crească copii, să îşi îngrijească părinţii în vârstă şi să se recalifice profesional fără să ajungă să trăiască de pe o zi pe alta. Pe lângă toate acestea, uniunile trebuie să se reinventeze pentru a-i ajuta pe angajaţi să aibă mai multe drepturi.

A cincea soluţie ar fi reglementarea mai strictă a monopolurilor de date. În cele din urmă, ar putea însemna că anumite companii de IT precum Alphabet sau Facebook va trebui să fie sparte în bucăţi, aşa cum s-a întâmplat şi în cazul industriilor petrolului şi telecomunicaţiilor în trecut. Înainte să se ajungă la asta, am putea lua în considerare şi schimbări mai puţin radicale. Companiilor mari de tehnologie le poate fi interzis doar să cumpere companii noi. Şi, aşa cum voi discuta în capitolul următor, cu siguranţă trebuie să ne gândim la nişte legi noi în ceea ce priveşte protecţia datelor şi proprietatea datelor.

Schimbări de genul acesta ar putea fi prea mici pentru a ne asigura că toţi cei din societate vor beneficia de pe urma Revoluţiei IA. Aşa cum s-a întâmplat în cazul inventării statului social modern la începutul revoluţiei industriale, s-ar putea să trebuiască să ne gândim la nişte măsuri mai radicale.

Venitul universal de bază, un venit garantat şi necondiţionat pentru toţi cetăţenii unei ţări, este o posibilitate. Cei de dreapta îl văd ca o modalitate de reducere a birocraţiei guvernamentale, în timp ce pentru cei de stânga este o metodă prin care se redistribuie bani celor săraci. Şi este o idee susţinută de mulţi din Silicon Valley.

Să primească toată lumea nişte bani în mod necondiţionat ar putea părea ceva radical, dar este de fapt o extensie a ceea ce facem deja. În multe ţări oamenii primesc educaţie la şcoală gratis. Şi în unele ţări primeşti servicii medicale gratuit. Simplul fapt că te naşti într-o ţară anume înseamnă că este suficient ca să primeşti servicii în valoare de mii de dolari pe an. Să pui bani direct în mâinile oamenilor ar putea părea ceva mai radical, dar facem asta de ceva timp, în mod implicit.

Swinging sixties aproape că au adus un venit de bază

în Statele Unite. În anul 1967, Martin Luther King Jr a scris: „Acum sunt convins că cea mai simplă abordare se va dovedi a fi cea mai eficientă – soluţia pentru sărăcie este eradicarea directă a acesteia, printr-o măsură care acum este discutată la nivel mondial: venitul garantat”[72]. Anul următor, tinerii au făcut proteste de stradă peste tot în lume pentru a cere un viitor mai bun, iar 1200 de economişti au redactat o scrisoare deschisă care a apărut pe prima pagină a ziarului *New York Times*, în care cereau un venit de bază pentru toată lumea. „Ţara nu îşi va îndeplini responsabilităţile până când fiecare individ din această naţiune nu va avea un venit asigurat care să nu reprezinte mai puţin decât definiţia sărăciei recunoscută oficial”, au scris ei. Preşedintele Richard Nixon a încercat chiar să pună în practică această idee. După câteva experimente cu venitul de bază universal, ce au avut loc în New Jersey, Pennsylvania, Iowa şi Carolina de Nord, Nixon a iniţiat un proiect de lege în anul 1969 pentru a pune capăt Războiului împotriva Sărăciei. Acesta ar fi garantat unei familii de patru persoane o sumă de 1600 de dolari pe an (echivalentul a 10.000 de dolari în prezent). Proiectul de lege a fost aprobat de Congres, dar a fost respins de Senat. Nixon

---

[72] Vezi Martin Luther King Jr, *Where De We Go from Here: Chaos or Community?*, Beacon Press, Boston, 1967.

a încercat din nou anul următor, cu acelaşi rezultat.

Noua ameninţare a automatizării a adus acum în prim-plan din nou ideea unui venit de bază universal. Un număr de experimente se desfăşoară în Olanda, Canada, Finlanda şi în alte locuri. Acestea sunt concepute astfel încât să ofere răspunsuri întrebărilor legate de fezabilitatea sa. Vor mai dori oamenii să muncească? Ce impact va avea asupra respectului de sine al oamenilor? Unii se tem că ar putea chiar să ducă la scăderea salariilor, întrucât angajatorii nu vor mai considera că este nevoie să plătească un salariu minim.

Una dintre problemele fundamentale ale tuturor acestor studii-pilot este scara la care se desfăşoară. Niciunul nu a fost universal, şi niciunul nu a durat suficient de mult pentru ca rezultatele sale să poată fi generalizate pentru o ţară întreagă şi pentru o generaţie întreagă de oameni. Cu toate acestea, rezultatele preliminare obţinute în urma acestor experimente (şi a celor anterioare) au fost pozitive. Oamenii nu muncesc mai puţin, sănătatea devine mai bună în aceste comunităţi, iar rezultatele din educaţie se îmbunătăţesc.

Dar rămân problemele majore – poate că cea mai mare o constituie costul unui venit de bază universal. În Statele Unite, să dai 18.000 de dolari pe an la 200 de milioane de adulţi ar costa 3,6 trilioane de dolari pe an.

E amuzant că aceasta este suma pe care o obţine anual bugetul federal american. Dar un venit de bază universal nu implică faptul că poţi elimina toate celelalte cheltuieli ale guvernului. Tot trebuie să se plătească străzile, şcolile, spitalele şi alte bunuri publice de care depind oamenii.

S-au propus şi alternative mai puţin radicale decât venitul de bază universal. Acestea includ un salariu minim pe economie mai mare, legislaţie mai bună privind sindicatele şi un drept al muncii mai puternic, transferul taxelor de la angajaţi către deţinătorii de capital şi finanţare guvernamentală pentru instruirea profesională şi reeducare. Aceste alternative au avantajul că necesită schimbări mai puţin radicale la nivelul societăţii. Dar încă nu ştim dacă aceste măsuri vor fi de ajuns pentru a face faţă schimbărilor ce ne pasc, chiar dacă ar fi aplicate simultan.

Renaştem din cenuşă

Companiile de tehnologie încep să-şi dea seama de responsabilităţile lor. De exemplu, Google a anunţat în anul 2017 că va investi un miliard de dolari de-a lungul următorilor cinci ani în organizaţii nonprofit care îi ajută pe oameni să se adapteze la schimbările care au loc în privinţa modalităţilor de angajare şi a tipului de

muncă. Aceasta este cea mai mare contribuţie filantropică a companiei făcută vreodată. Dar dacă presupunem că profiturile sale vor continua să crească în ritmul din prezent, această iniţiativă va costa Google mai puţin de 3% din profiturile sale obţinute pe parcursul celor cinci ani cât va dura proiectul. Ar contribui mult mai mult la societate dacă ar plăti impozit pe profit, cum fac alte companii.

Un alt exemplu din anul 2017: Facebook a anunţat că nu va mai raporta în Irlanda toate veniturile sale din reclame câştigate din Europa, ci va începe să le declare în ţările în care sunt câştigate. Criticii au sugerat că această mişcare nu va avea ca rezultat faptul că Facebook va plăti mai multe taxe. Cu toate acestea, ar putea însemna începutul sfârşitului evaziunii fiscale agresive a companiilor de tehnologie.

Dar poate că cea mai promiţătoare soluţie nu vine din partea unei companii, ci a unui guvern. De-a lungul ultimilor 60 de ani, guvernele din Costa Rica nu s-au concentrat pe creşterea economică, ci pe a oferi accesul tuturor cetăţenilor săi la servicii sociale de calitate înaltă şi, în acelaşi timp, pe protejarea mediului. Abolirea armatei şi folosirea resurselor astfel eliberate pentru spitale, şcoli şi pensii au transformat Costa Rica într-un exemplu pentru America Latină.

În anul 2016, Costa Rica a cheltuit 6,9% din bugetul

său pe educaţie, în comparaţie cu media globală de doar 4,4%. Banii cheltuiţi pe serviciile de sănătate au crescut cu în jur de 50% în ultimii 20 de ani, ajungând la 9% din produsul intern brut. Şi peste 70% din această sumă reprezintă finanţări de stat. Costa Rica plănuieşte să ajungă neutră din punctul de vedere al emisiilor de carbon până în anul 2021. Peste 98% din electricitatea sa este deja obţinută din resursele regenerabile. Speranţa sa de viaţă de 79,6 ani a depăşit-o pe cea din Statele Unite ale Americii. Costa Rica a devenit un loc mai bun pentru cetăţenii săi cheltuind bani pe sănătate, educaţie şi mediu. Să sperăm că şi alte ţări observă acest lucru.

Nu va fi uşor să ne descurcăm cu inegalitatea în creştere. Va fi nevoie de îndrăzneală şi de viziune. Cel mai sigur va fi nevoie ca profiturile pe care le aduce IA companiilor tehnologice mari să fie redistribuite mai larg.

În cadrul World Economic Forum din anul 2018, Sundar Pichai, CEO al Google, a declarat că este bucuros că Google va plăti mai multe taxe şi a cerut ca sistemul global de taxe actual să fie modificat. Viitorul pare deja destul de evident. Până în anul 2062, companii precum Google ar trebui într-adevăr să plătească mai multe taxe, şi aceste taxe trebuie să fie distribuite global.

0008
Sfârşitul spaţiului privat

Spaţiul privat este încă o valoare umană care se află în pericol. Vint Cerf este „evanghelistul-şef al internetului” la Google şi unul dintre arhitecţii internetului. În anul 2013 a spus Comisiei Federale pentru Comerţ că „spaţiul privat ar putea de fapt să fie o anomalie”. A justificat această declaraţie îndrăzneaţă cu observaţia că „spaţiul privat este ceva ce a apărut în urma exploziei urbane rezultate datorită revoluţiei industriale”.

Există o măsură de adevăr în ceea ce susţine el. În epoca medievală, exista mult mai puţină intimitate. Majoritatea oamenilor nu-şi puteau permite să locuiască în case cu spaţii separate şi cu dormitoare. Revoluţia Industrială ne-a ridicat standardele de viaţă şi au făcut ca anumite forme de spaţiu privat să fie posibile. Dar spaţiul privat înseamnă mai mult decât propria cameră. Înseamnă dreptul la intimitate pentru a discuta despre schimbări politice şi pentru a gândi planuri periculoase. Dreptul de a vota anonim. De a practica religia pe care o alegem. De a trăi stilul de viaţă pe care ni-l dorim. Sunt multe alte aspecte ale vieţii private pe care am ajuns să ne aşteptăm să le avem.

Totuşi, până în anul 2062 nu vom mai considera ca de la sine înţelese unele dintre aceste aspecte. Foarte puţine lucruri vor fi scutite de controlul atent al unei inteligenţe artificiale foarte eficiente. Atunci, o întrebare vitală este cum vom opri IA să nu compromită dreptul la o viaţă privată, aşa cum a prezis George Orwell în celebrul său roman distopic *1984.*

Noul petrol

Clive Humby, matematician britanic şi inventatorul din spatele inovatorului Clubcard al companiei Tesco, este considerat a fi prima persoană la nivel mondial care a comparat datele cu petrolul[73]. Bineînţeles, petrolul a fost combustibilul care a alimentat cea mai mare parte a revoluţiei industriale, la propriu şi la figurat. În anul 2006, Humby a spus: „Datele sunt noul petrol. Ţiţeiul e valoros, dar dacă nu e rafinat, nu prea poate fi folosit. Trebuie să fie prelucrat în gaze, plastic, chimicale etc. Pentru a se crea o entitate valoroasă ce conduce o activitate profitabilă; în mod similar, datele trebuie să fie descompuse şi analizate pentru a avea valoare".

În anul 2013, Abhishek Mehta, directorul companiei

[73] Discurs la ANA Senior Marketers's Summit 2006 la Şcoala de Management Kellogg, Universitatea Northwestern.

Tresata, a făcut următoarele observaţii:

Aşa cum petrolul a fost o resursă naturală care a alimentat ultima revoluţie industrială, datele vor reprezenta resursele naturale pentru această nouă revoluţie industrială. Datele constituie materia primă şi lubrifiantul de bază nu doar pentru toate modelele economice din jurul fiecărei industrii, ci şi pentru modelele socio-economice.

IA a fost şi va continua să fie un consumator major de date. Metodele de învăţare automată precum învăţarea profundă necesită în prezent milioane, dacă nu chiar miliarde de exemple pentru antrenament. Dacă datele sunt noul petrol, învăţarea mecanică este noua rafinărie a acestor baze de date imense. Până în anul 2062 învăţarea mecanică va ajunge să fie, fără îndoială, mult mai asemănătoare cu învăţarea umană, şi va necesita mai puţine exemple. Cu toate acestea, datele vor fi în continuare centrul succesului IA.

Analogia dintre date şi petrol nu ar trebui luată chiar atât de literal, întrucât există câteva diferenţe fundamentale. Petrolul este o resursă preţioasă şi limitată. Datele nu sunt nici preţioase, nici rare. Petrolul poate fi utilizat o singură dată. Datele pot fi refolosite fără nicio limită. Spre deosebire de petrol,

datele pot fi folosite adesea pentru a genera mai multe date. Piero Scaruffi, autorul cărţii A History of Silicon Valley, a scris: „Diferenţa dintre petrol şi date este că produsul petrolului nu generează mai mult petrol (din păcate), în timp ce produsul datelor (autovehiculele care se conduc singure, dronele, calculatoarele portabile etc.) va genera mai multe date (unde conduci de obicei, cât de repede/ bine conduci, cine e cu tine etc.)”[74].

Dar poate că cea mai mare diferenţă dintre petrol şi date priveşte proprietatea asupra lor. Ţările au declarat rapid dreptul de proprietate asupra petrolului de sub picioarele noastre şi de sub mări. Dar o mare parte din date sunt deţinute în privat. Câteva monopoluri de date private, în special Google şi Facebook, deţin din ce în ce mai multe date ale noastre. Şi în timp ce obţin bani de pe urma lor, noi – producătorii de date – ne bucurăm de foarte puţin din valoarea lor. În plus, toate aceste date ne pun în pericol intimitatea.

IA e cu ochii pe tine

În anul 2062 va fi o mare provocare să îţi asiguri spaţiul privat de care ai nevoie. The Economist a

[74] În *Humankind 2.0,* o carte despre progres; vezi scarufifi.com/singular/bigdata.html

estimat că doar Facebook a scanat, a înregistrat şi poate să recunoască 1,2 miliarde de feţe. Gândiţi-vă la asta pentru o clipă. În câteva decenii, cel mai probabil va exista o bază de date undeva, al cărei scop va fi să cuprindă toate feţele.

Nici măcar animalele noastre de companie nu sunt în siguranţă: în octombrie 2017, Google Photos a început să recunoască şi să eticheteze câini şi pisici. Şi când Google îţi vede pudelul, va fi pregătit să te caute prin apropiere. Aşa că data viitoare când etichetezi un prieten sau un animal de companie pe o reţea de socializare, ţine minte că dezvălui nu doar identitatea animăluţului, ci şi a prietenului tău.

Silicon Valley ar putea avea idee nu doar despre cine eşti, dar şi pe cine vei vota. Sau despre alte informaţii private, cum ar fi sexualitatea ta. În 2017, o echipă de la Universitatea Stanford a demonstrat că poate prezice cum vei vota, folosindu-se pur şi simplu de imaginile de pe Google Street View[75]. Până şi un singur like de pe Facebook poate să ofere informaţii suficiente pentru a

[75] Vezi Timnit Gebru, Jonathan Krause, Yilun Wang, Duyun Chen, Jia Deng, Erez Lieberman Aiden & Li Fei-Fei, „Using Deep Learning and Google Street View to Estimate the Demographic Makeup of Neighborhoods Across the United States" *Proceedings of the National Academy of Sciences,* vol. 114, nr. 50, 2017, pp. 13108-13113.

se prezice cum vei vota[76]. O altă echipă de la Universitatea Stanford a făcut o declaraţie controversată cum că au antrenat un algoritm de învăţare mecanică să facă diferenţa dintre feţele de persoane homosexuale şi cele de persoane heterosexuale.[77]

Par să existe foarte puţine limite în ceea ce priveşte datele pe care companiile tehnologice îşi doresc să le colecteze. Bruce Schneier, expert în confidenţialitate şi

---

[76] Vezi Jakob Baek Kristensen, Thomas Albrechtsen, Emil Dahl-Nielsen, Michael Jensen, Magnus Skovrind & Tobias Bornakke (2017), „Parsimonious Data: How a Single Facebook Like Predicts Voting Behavior in Multiparty Systems", *PLoS One*, vol. 12, nr. 9, eO184562.

[77] Vezi Yilun Wang & Michal Kosinski, „Deep Neural Networks Are More Accurate than Humans at Detecting Sexual Orientation from Facial Images", *Journal of Personality and Social Psychology*, vol. 114, nr. 2, 2018, pp. 246-257. Sunt multe lucruri de criticat la acest proiect de cercetare. Studiul nu a urmărit un eşantion reprezentativ. Nu a inclus decât oameni albi care locuiesc în Statele Unite, cu vârste cuprinse între 18 şi 40 de ani. A considerat că orice persoană este ori homosexuală, ori heterosexuală. Datele pentru antrenare şi pentru test au folosit un număr egal de persoane homosexuale şi heterosexuale, în timp ce în realitate numai 7% din persoanele din acest grup de vârstă sunt homosexuale. Studiul a declarat că algoritmul poate să facă distincţia dintre imaginea unui bărbat homosexual şi cea a unui bărbat heterosexual în 81 % din cazuri. Cu toate acestea, acurateţea în cazul unui set de teste echilibrat din punct de vedere demografic ar sta mult mai prost. Şi, la urma urmei, care sunt posibilele beneficii ale unui astfel de studiu? Dezvoltarea unui software care să recunoască persoanele homosexuale prezintă un risc imens. Sunt zeci de ţări care au pedeapsa cu moartea pentru homosexualitate. Oamenii se îmbracă şi se aranjează astfel încât să se conformeze cu diverse stereotipuri culturale şi sexuale. Fără îndoială, algoritmul de învăţare mecanică şi-a dat seama de aceşti factori. Nu avem nevoie de un software care să ne spună asta.

securitate, a observat că „supravegherea este modelul de afaceri al internetului”[78]. Al Gore s-a exprimat şi mai succint, declarând-o „economia hărţuitoare”[79]. De ce s-a gândit Google că ar fi cazul să îţi urmărească telefonul Android când îţi dezactivai localizarea şi chiar şi când scoteai cardul SIM[80]? De ce s-a gândit Uber că ar fi în regulă să îţi urmărească locaţia la cinci minute după ce s-a terminat cursa[81]? De ce s-a gândit *Pokémon Go* că ar trebui să aibă acces la contul tău întreg de Google de pe iOS, inclusiv la e-mail-ul tău şi la istoricul de căutări?

Desigur, companiile tehnologice nu sunt singurele care încep să invadeze spaţiul privat al oamenilor. Statele de asemenea adoptă noi tehnologii pentru a se băga în vieţile noastre. Statul american Delaware, spre exemplu, pune camere inteligente în maşinile poliţiei ca să detecteze vehiculele în care se află evadaţi, copii răpiţi sau bătrâni dispăruţi. Astfel de utilizări ale tehnologiei ar putea părea neproblematice în

---

[78] Bruce Schneier, „Stalker economy; Here to Stay“, *CNN.com*, 25 noiembrie 2013.

[79] „Guvernele - ele sunt o ameninţare; firmele de afaceri colectează şi ele mai multe informaţii decât este necesar. Avem acum o economie hărţuitoare în care firmele află totul despre tine. " Al Gore, în cadrul Conferinţei Southland, 10 iunie 2014.

[80] Shannon Liao, „Google Admits It Tracked User Location Data Even Ehen the Setting Was Turned Off", *The Verge*, 21 noiembrie 2017.

[81] Amar Toor, „Uber Will No Longer Track Your Location After Your Ride Is Over", *The Verge*, 29 august 2013.

momentul de faţă, dar ce se întâmplă când un stat va începe să folosească aceeaşi tehnologie pentru a găsi activiştii politici sau refugiaţii? În China, poliţia din Zhengzhou a început să testeze ochelari cu recunoaştere facială. Aceşti ochelari pot să proceseze 100.000 de feţe pe secundă. Nu ai cum să te ascunzi într-o mulţime de protestatari când este disponibilă o astfel de tehnologie.

Date interconectate

Chiar dacă o companie s-a comportat responsabil şi a anonimizat datele pe care le-a colectat, tot poate să dea greş. Una dintre probleme este conectarea datelor din diverse surse. Luat separat, s-ar putea ca un set de date să nu dea nimic privat în vileag. Dar atunci când două sau mai multe seturi de date sunt combinate, intimitatea ta poate fi ameninţată. Provocarea lansată de Netflix reprezintă un exemplu bun.

În anul 2006, Netflix a lansat o competiţie de un milion de dolari pentru a concepe un sistem mai bun de recomandare a filmelor. Compania a publicat 100.480.507 de ratinguri ce au fost oferite de 480.189 de utilizatori unui număr de 17.770 de filme diferite. Pentru a proteja intimitatea utilizatorilor, Netflix a avut grijă ca datele să fie anonime, eliminând detaliile

personale şi înlocuind numele cu numere la întâmplare. Dar asta nu a funcţionat. Cercetătorii de la Universitatea din Texas din Austin au reuşit să identifice utilizatorii din setul de date al Netflixului, comparând datele respective cu punctajele şi marcajele de timp din baza de date disponibilă public Internet Movie Database (IMDb). Se pare că, dacă scoţi din setul de date cele mai apreciate filme care ne plac tuturor, poţi să identifici cu mare uşurinţă pe cineva din combinaţia filmelor mai puţin populare care îi plac.

Netflix ar fi putut să îngreuneze recunoaşterea utilizatorilor prin eliminarea unora dintre date, prin modificarea marcajului de timp sau prin introducerea unor erori deliberate în date. Dar până şi cu date parţiale, cu date care au fost modificate sau cu date cu erori în ele, tot nu a fost chiar atât de greu ca unii dintre utilizatori să fie identificaţi. A fost necesară doar o cantitate mică de date non-anonime pentru a deconspira cel puţin o parte din baza de date mult mai mare, dar anonimă a Netflixului.

Guvernele sunt foarte conştiente de puterea datelor interconectate. În anul 2015, *Data Retention Act* a fost promulgat în Australia. Acest lucru a introdus o obligaţie legală pentru companiile de telecomunicaţii să reţină anumite tipuri de metadate. Pentru apelurile telefonice, aceste date includ numărul apelantului,

locaţia dispozitivului şi numărul unic de identificare al telefonului mobil. Pentru e-mail-uri, sunt incluse adresa de e-mail a expeditorului, mărimea şi data calendaristică a mesajului trimis.

Nu se pot face prea multe doar cu metadatele, deoarece nu includ conţinutul apelului sau al e-mail-ului. Dar dacă sunt juxtapuse cu alte date, autorităţile pot descoperi mai multe lucruri – cum ar fi cu cine te asociezi şi ce faci.

Spaţiul privat offline

Chiar şi atunci când suntem offline, suntem urmăriţi din ce în ce mai mult. O companie privată din Statele Unite, Vigilant Solutions, are o bază de date de peste 2,2 miliarde de fotografii cu plăcuţe cu numere şi locaţii. În fiecare lună, compania creează şi stochează permanent alte în jur de 80 de milioane de intrări. Vigilant Solutions vinde aceste date câtorva mii de secţii de poliţie din toată ţara care vor să urmărească oameni. În prezent are un contract cu Departamentul Securităţii Naţionale, pentru a îi oferi urmărirea în timp real a plăcuţelor cu numere de înmatriculare.

În mod nesurprinzător, cumpărăturile tale sunt şi ele urmărite cu atenţie. Google, spre exemplu, adună date de marketing de la Adwords Google Analytics şi

Doubleclick Search, împreună cu date de localizare preluate de pe telefoane mobile, pentru a vedea când oamenii vizitează site-ul unui magazin. În fiecare an, urmăresc miliarde de vizite la magazine după ce utilizatorii au dat click pe anumite reclame. Şi Google a început să conecteze aceste vizite la magazin cu datele de cumpărare. Parteneriatele lui Google cu „terţe părţi" deja înregistrează aproximativ 70% din toate tranzacţiile prin carduri de credit şi de debit din Statele Unite. Până în anul 2062, ne putem aştepta ca firme precum Google să urmărească toate cumpărăturile noastre, atât online, cât şi offline. Va exista o bază de date care va înregistra fiecare dolar pe care îl cheltui.

Şi odată ce oraşele vor deveni „mai inteligente" şi vor colecta, analiza şi procesa datele despre locuitorii lor, supravegherea offline s-ar putea să ajungă să fie şi mai indiscretă. Chiar nu există nicio modalitate de a opta să nu fii inclus în astfel de colectări de date. Toată lumea care îşi vede de treaba sa va fi înregistrată. De exemplu, în anul 2013 s-a descoperit că noile „coşuri de gunoi inteligente" din Londra urmăreau telefoanele mobile ale cetăţenilor[82]. Kaveh Memari, directorul companiei care a produs coşurile inteligente, a vorbit foarte deschis despre intenţiile firmei sale: „Din punctul

[82] Matt Warman, „Bins that Track Mobiles Banned by City of London Corporation", *The Telegraph*, 12 august 2013.

nostru de vedere, sunt accesibile tuturor, toată lumea poate să cumpere acele date. Londra este locul cel mai supravegheat din lume... Atât timp cât nu includem un nume şi o adresă, este legal”[83].

Argumentul lui are o logică stranie: pentru că alţii urmăresc oameni, putem face şi noi la fel. În urma unui scandal public, au fost forţaţi să oprească coşurile inteligente din a urmări telefoanele oamenilor – deocamdată. Dar până în anul 2062, nu doar coşurile de gunoi vor fi cele care vor urmări oameni. Întregul oraş va fi cu ochii pe tine.

Big Brother la tine acasă

Supravegherea nu se va opri atunci când vei ajunge acasă. Deja începem să vedem asta prin difuzoarele inteligente ca Alexa de la Amazon şi Google Home. Deşi acestea „se trezesc” numai atunci când vorbeşti cu ele, motivul pentru care sunt capabile să facă asta este că ascultă *mereu.* Şi odată ce s-au trezit, îţi procesează vorbele nu pe dispozitiv, ci în cloud, care se găseşte în serverele Google şi Amazon. Google şi Amazon înregistrează o conversaţie doar atunci când dispozitivul este pornit. Dar cercetătorii au descoperit

[83] Siraj Datoo, „This Recycling Bin Is Following You" *Quartz*, 8 august 2013.

cum este posibil să le accesezi ilegal şi să înregistrezi conversaţii tot timpul. Acest lucru transformă difuzorul într-un microfon virtual. Există un motiv pentru care Edward Snowden îşi acoperă camera de la laptop cu bandă şi îşi pune telefonul în frigider ca să blocheze semnalul.

Poliţiştii din Arkansas au solicitat înregistrările unui client Alexa în legătură cu omuciderea care a avut loc la el acasă în anul 2015. Amazon a refuzat să le înmâneze înregistrările, dar înainte ca problema să fie discutată la tribunal, avocatul clientului a fost de acord să înmâneze datele. Până în 2062, ne putem aştepta ca tot soiul de asistenţi personali digitali să depună mărturii în cazurile legale.

Nu doar difuzoarele inteligente ne vor invada intimitatea în propria casă. În viitor, internetul va evolua şi va deveni Internet of Things, „internetul obiectelor", la care vor fi conectate toate dispozitivele din casele noastre: televizoarele, frigiderele, aparatele de prăjit pâine, luminile, până şi ghivecele cu flori. Cele mai multe nu vor avea ecran sau tastatură, dar vor avea interfeţe cu comandă vocală. Aşadar, vor avea microfoane care vor asculta în permanenţă comenzi. În cartea *1984* a lui Orwell, guvernul era cel care asculta în casele oamenilor. În realitate, deja plătim companii private ca să pună în casele noastre dispozitive care pot

asculta orice conversaţie.

Spaţiul privat analog

Unii susţin că lupta pentru spaţiul nostru privat digital este deja pierdută. Deja am oferit prea multe informaţii private companiilor Facebook şi Google, Amazonului şi altora. Dar în curând vom renunţa şi la viaţa noastră privată analogă. Problema este că ne conectăm singuri la ceasurile inteligente, la brăţările de monitorizare biomedicală şi la alte dispozitive care ne monitorizează. În felul ăsta ne divulgăm singuri locaţia geografică, pulsul, presiunea arterială – şi în curând vom mai adăuga şi alte funcţii vitale.

Beneficiile sunt evidente. Aplicaţia de învăţare automată Deepheart pentru Apple Watch poate să detecteze fibrilaţia atrială, hipertensiunea şi apneea din timpul somnului. Poate chiar să se folosească de bătăile inimii tale ca să prezică un început de diabet cu o acurateţe remarcabilă de 85%. Deci avem motive să sperăm că IA ne va face mai sănătoşi. Dar există şi riscuri. Dacă compania ta de sănătate îţi măreşte costul asigurării în fiecare săptămână când sari peste sală? Sau dacă angajatorul tău te penalizează pentru că lucrezi prea încet? Sau dacă un comerciant îţi arată reclame care îţi fac inima să-ţi bată mai repede?

În lumea digitală putem minţi cu uşurinţă cu privire la cine suntem. Putem să ne prefacem că suntem ceea ce nu suntem. Putem să ne conectăm anonim. Dar este mult mai greu să minţi în legătură cu sinele tău analog. Avem foarte puţină putere de control directă asupra ritmului în care ne bate inima sau asupra felului în care ni se dilată pupilele. Imaginaţi-vă ce ar putea face un partid politic dacă ar avea acces la pulsul tuturor! Iar noi dăm aceste date analoge unor companii private.

De exemplu, atunci când te înregistrezi să foloseşti un dispozitiv Fitbit, acesta colectează o mulţime de date analoge despre tine, pe care le stochează în serverele sale: numărul de paşi pe care i-ai făcut astăzi, distanţa pe care ai parcurs-o, caloriile pe care le-ai ars, greutatea ta actuală, pulsul şi locaţia, precum şi cele mai apropiate puncte de acces Wifi, coordonatele releelor de telefonie mobilă, calculatoarele pe care le foloseşti şi paginile web pe care le accesezi. Fitbit poate să-şi dea seama de foarte multe lucruri despre tine cu ajutorul acestor date.

Alt exemplu: atunci când trimiţi o parte din saliva ta la AncestryDNA pentru o testare genetică, trebuie să fii de acord cu faptul că le acorzi „o licenţă gratuită, mondială, sublicenţiabilă, transferabilă să găzduiască, să transfere, să proceseze, să analizeze, să distribuie şi să comunice informaţiile tale genetice cu scopul de a-ţi

oferi produse şi servicii, de a face cercetare cu privire la o descendenţă, de a îmbunătăţi experienţa utilizatorilor şi de a crea şi a oferi produse şi servicii personalizate". Iar dacă se întâmplă ca AncestryDNA să îţi folosească ADN-ul pentru a găsi leacul pentru o boală genetică rară pe care o ai tu, este legal să te facă să plăteşti pentru acel leac. Termenii şi condiţiile de la AncestryDNA îţi spun destul de clar că „nu obţii niciun drept în nicio cercetare sau niciun produs comercial care poate fi dezvoltat de Ancestry pe baza informaţiilor tale genetice". Aceste condiţii au fost chiar şi mai dezavantajoase la un moment dat. Înainte de protestele de pe reţelele de socializare, AncestryDNA cerea utilizatorilor să-i acorde o licenţă gratuită „perpetuă" Odată ce îţi obţineau datele, nu exista nicio cale prin care puteai să ţi le retragi. Acum poţi măcar să le ceri să îţi şteargă datele şi să nu le mai folosească.

Spre deosebire de datele medicale colectate de doctori şi de spitale, datele analoge colectate de Fitbit şi AncestryDNA nu sunt protejate de nicio legislaţie pacient/doctor sau legislaţie ce ţine de intimitatea pacientului. O companie ca Fitbit sau AncestryDNA poate să facă orice doreşte cu aceste date. Fitbit poate detecta cine întreţine relaţii sexuale şi să încerce să îi

vândă Viagra[84]. AncestryDNA poate determina că eşti supus riscului de a te îmbolnăvi de Alzheimer şi îţi va vinde detaliile unui azil local.

Ochi non-umani

Unul dintre argumentele pentru a permite acestor tehnologii să continue să opereze în aceleaşi moduri este că nicio pereche de ochi umani nu se uită peste datele respective. Până în anul 2017, serverele Google îţi citeau e-mail-urile primite şi îţi foloseau informaţiile pentru a oferi reclame personalizate. Bineînţeles, nu era o persoană adevărată cea care îţi citea e-mail-urile, ci un algoritm. Cu toate acestea, tot poate părea ciudat. Directorul executiv de la Google, Eric Schmidt, a declarat: „Politica Google este să meargă până la limita la care ceva poate fi considerat ciudat şi să nu o depăşească”[85]. Dar decizia recentă a companiei de a înceta să citească e-mail-uri sugerează că poate şi-a dat seama că a depăşit deja acea linie[86].

În trecut, Google s-a apărat în ceea ce priveşte citirea

---

[84] În anul 2011, activitatea sexuală a multor utilizatori FitBit putea fi găsită în rezultatele de pe Google. Acum este interzisă divulgarea acestor informaţii.

[85] Eric Schmidt, Washington Ideas Forum, octombrie 2010.

[86] Google nu mai citeşte e-mail-urile pentru a oferi reclame personalizate. Dar continuă să le citească cu alte scopuri, cum ar fi adăugarea unor notiţe în calendar sau sugerarea unor răspunsuri automate.

e-mail-urilor utilizatorilor folosind acest argument: „La fel cum expeditorul unei scrisori adresate unui coleg de muncă nu poate fi surprins că asistentul destinatarului deschide scrisoarea, oamenii care folosesc corespondenţa electronică în ziua de azi nu pot să fie surprinşi atunci când comunicarea lor este procesată de compania de SEC (servicii electronice de comunicaţii) pe parcursul unei livrări”[87]. Dar argumentul lui Google n-are nicio logică. Nu te-ai aştepta ca un angajat la poştă să îţi citească scrisoarea atunci când ţi-o livrează. Şi ai fi puţin dezamăgit dacă ar citi dincolo de adresa care este menţionată pe felicitarea ta. Aşa că ar trebui să fim surprinşi şi dezamăgiţi că şi *conţinutul* e-mail-urilor noastre e citit în timp ce ele sunt în curs de a fi livrate. De vreme ce IA devin din ce în ce mai bune, ar trebui să fim din ce în ce mai îngrijoraţi în privinţa ochilor lor non-umani care ne citesc corespondenţa.

Apple cel bun

O singură companie a încercat să se diferenţieze de restul în ceea ce priveşte viaţa privată a oamenilor. Declaraţia privind viaţa privată a celor de la Apple spune destul de clar:

---

[87] Gregory S. McNeal, „It's Not a Surprise that Gmail Users Have No Reasonable Expectation of Privacy“, *Forbes*, 20 august 2013.

Produsele Apple sunt concepute pentru a face lucruri extraordinare. Şi sunt concepute pentru a-ţi proteja intimitatea.

La Apple, noi credem că viaţa privată este un drept fundamental al omului.

Şi o mare parte din informaţiile tale personale – informaţii pe care ai dreptul să le păstrezi pentru tine – se găsesc în dispozitivele Apple.

Pulsul tău după ce ai alergat. Ce ştiri citeşti mai întâi.

De unde ţi-ai cumpărat ultima cafea. Ce site-uri web accesezi. Pe cine suni, cui dai mesaj şi cui e-mail.

Fiecare produs Apple este creat de la zero pentru a-ţi proteja informaţiile. Şi pentru a-ţi da puterea de a alege ce împarţi şi cu cine.

Am demonstrat de-a lungul timpului că experienţele minunate nu trebuie să se petreacă în detrimentul intimităţii şi al securităţii tale. În schimb, le pot susţine.

În sprijinul acestor declaraţii ferme, Apple a rezistat mai multor zeci de solicitări din partea guvernului american care a cerut să i se ofere acces la datele stocate de dispozitivele Apple. Chiar dacă unul dintre aceste dispozitive îi aparţinea unui terorist. Pentru asta, îi aplaud.

Dar când vine vorba de a alege între intimitatea

oamenilor şi profituri, Apple a acţionat mai puţin virtuos. În luna februarie, anul 2018, pentru a se conforma noilor legi chinezeşti, toate datele care aparţin utilizatorilor icloud de pe teritoriul Chinei au fost mutate pe serverele unei companii deţinute de guvernul Chinei. Termenii şi condiţiile serviciului Apple au fost modificate pentru a oferi permisiunea acestei companii de a accesa datele.

Nu a fost prima dată când Apple s-a plecat în faţa guvernului chinez. În 2017, software-ul VPN care le permite utilizatorilor să acceseze site-uri web interzise în China a fost scos din App Store. Este clar că Apple va ceda în faţa presiunii din partea guvernului Chinei pentru a se menţine pe această piaţă extrem de valoroasă. Prin comparaţie, Google a rezistat multor încercări din partea Chinei de a le cenzura rezultatele de căutare. Ca rezultat, Google a fost scos în totalitate de pe piaţa chineză. Spre deosebire de Apple, Google ar trebui să fie aplaudat pentru că a înfruntat China şi pentru că şi-a pus principiile mai presus de profituri.

Sistemul creditelor sociale

Probabil cea mai îngrijorătoare predicţie legată de viaţa noastră privată în anul 2062 este sistemul creditelor sociale care este dezvoltat în China. Până în

anul 2020, China plănuieşte să colecţioneze toate informaţiile valabile online despre companii şi cetăţeni într-un singur loc, după care să desemneze un scor pe baza lor, care va măsura cât de „demn de încredere" este cineva. Obiectivul este de a „le oferi beneficii celor demni de încredere şi de a-i disciplina pe cei nedemni de încredere... (astfel încât) integritatea să devină o valoare socială răspândită"[88]. Documentele oficiale legate de această schemă oferă prea puţine detalii concrete, dar au sugerat că „cei nedemni de încredere" vor fi pedepsiţi prin restricţii în privinţa angajării, a călătoriilor, a locuinţei şi a tranzacţiilor bancare.

Implementările-pilot care sunt în desfăşurare acum nu prea au risipit îngrijorările legate de acest sistem. Unul dintre cele mai vizibile proiecte-pilot este Sesame Credit, un sistem cu notare prin credite condus de Alibaba. În timp ce Amazon are 310 milioane de conturi de clienţi, Alibaba este chiar şi mai mare, cu aproape o jumătate de miliard de utilizatori lunar. Procesează peste 11% din vânzările cu amănuntul din China. Detaliile legate de felul în care este calculat Sesame Credit sunt secrete. Dar ia în considerare cinci factori: achiziţiile făcute prin intermediul platformei

[88] Stefan Brehm & Nicholas Loubere, „Chinas Dystopian Social Credit System is a Harbinger of the Global Age of the Algorithm", *The Conversation*, 15 ianuarie 2018.

Alipay de plată online şi pentru dispozitive mobile ale companiei Alibaba, informaţii personale, informaţii privind plata la timp a facturilor, plata la timp a cardului de credit şi prietenii unei persoane. E mai bine să nu ai prieteni care nu sunt demni de încredere!

Până acum, sistemul a folosit mai mult recompense decât pedepse. De exemplu, cei care au un scor bun pot să facă rezervări pentru camere de hotel şi să închirieze biciclete fără avans. O vreme au putut să acceseze linia cu prioritate din aeroportul din Beijing. Iar cel mai mare serviciu online de întâlniri romantice din China îi promovează pe cei care au scoruri bune. Dar China a anunţat că începând cu luna mai a anului 2018 oamenii cu credite scăzute nu vor mai putea circula cu trenul sau cu avionul. Potenţialul ca acest sistem să fie folosit în moduri greşite este imens, iar China nu este o ţară cu un istoric bun în ceea ce priveşte drepturile omului[89].

Post-Snowden

China nu este singura ţară pentru care ar trebui să ne facem griji. În anul 2013, Edward Snowden a dat în

[89] China ducea lipsă de un birou de credit precum cel din Statele Unite, aşa că pare rezonabil că guvernul chinez a dorit să aibă un sistem ca acesta pentru a combate, spre exemplu, corupţia omniprezentă. Dar apoi sistemul s-a transformat în ceva cu mult mai periculos.

vileag existenţa a numeroase programe de supraveghere globale, conduse de agenţii de informaţii din Statele Unite, Australia, Canada, Noua Zeelandă şi Marea Britanie. E-mail-uri, mesaje directe şi conversaţiile de pe telefonia fixă şi de pe telefoanele mobile au fost ascultate toate. Scopul a fost evident: „Colectează tot", „Procesează tot" şi „Exploatează tot". Şi nu doar inamicii noştri au fost ţinta; cetăţenii care respectau legea ai ţărilor respective au fost şi ei prinşi în această capcană imensă. Nu-i de mirare că lumea a fost furioasă. Astfel de supraveghere ascunsă încalcă *Al Patrulea Amendament* din Constituţia SUA, care interzice percheziţiile şi confiscările nerezonabile şi pretinde un mandat de percheziţie emis de un judecător pe baza unei suspiciuni întemeiate.

Având în vedere toate acestea, tot nu pot să înţeleg de ce atât de mulţi oameni au fost surprinşi că e-mail-urile lor erau citite. E-mail-ul este tipul de comunicare cel mai uşor de interceptat.

Poate fi citit direct de maşini. Spre deosebire de conversaţiile telefonice, nu e nevoie să fie transcris. Monitorizarea e-mail-urilor necriptate este un lucru prea uşor şi prea tentant pentru state.

Din nefericire, IA va face ca supravegherea cetăţenilor să devină şi mai uşoară. Algoritmii de recunoaştere vocală pot să asculte simultan milioane de

apeluri telefonice. Algoritmii vizuali ai calculatorului pot să privească simultan milioane de camere de supraveghere. Iar algoritmii care procesează limbajul natural pot să citească simultan milioane de e-mail-uri.

Europa ca lider

Europa oferă ceva speranţă pentru viaţa privată în anul 2062. În mai 2018, *General Data Protection Regulation* (GDPR) a fost implementat în Uniunea Europeană. Obiectivul principal al actului este să ofere cetăţenilor europeni controlul asupra datelor lor personale. Le oferă nişte drepturi de bază asupra datelor lor, cum ar fi dreptul de a le accesa şi de a le şterge.

Poate că cel mai relevant aspect al discuţiei despre viaţa privată ţine de dreptul de a primi o explicaţie. Când o IA ia decizii automate, actul spune că avem dreptul la „informaţii relevante despre logica folosită, precum şi despre semnificaţia şi consecinţele anticipate ale unui astfel de proces”. Încă nu ştim cum vor interpreta instanţele acest lucru. Dar ar putea asigura faptul că cetăţenii europeni primesc explicaţii utile legate de felul în care un algoritm ia decizii, precum şi dreptul de a nu fi vizaţi de astfel de decizii.

Pentru a obţine acordul de a colecta şi de a folosi

datele tale, GDPR le interzice companiilor să folosească termeni şi condiţii lungi, ilizibile, pline de jargon juridic. Consimţământul trebuie să fie clar şi să privească explicit anumite chestiuni, care să fie prezentate într-o formă inteligibilă şi uşor accesibilă, folosind un limbaj clar şi simplu. De asemenea, trebuie să fie la fel de uşor să îţi retragi acordul pe cât este să ţi-l dai.

Stimulentele care determină companiile să respecte actul sunt foarte mari. Organizaţiile care încalcă GDPR pot să fie amendate cu până la 4% din cifra lor de afaceri anuală sau 20 de milioane de euro (suma care este mai mare din cele două). Încă nu am văzut care este impactul acestor reglementări, dar pare a fi un prim pas către păstrarea caracterului privat al datelor personale.

Proprietarii datelor

GDPR este doar începutul în ceea ce priveşte schimbările care sunt necesare pentru a ne păstra vieţile private. Cum se poate ca o companie precum Facebook să nu genereze aproape niciun conţinut, dar să deţină tot conţinutul? Nu ar fi lumea un loc mai drept dacă am fi proprietarii datelor noastre şi dacă am avea dreptul să spunem cine le poate folosi? Şi dacă

acesta ar fi un drept, şi nu o „opţiune de retragere”?

La Congresul Mobile World din anul 2018, directorul în tehnologie de la IBM Watson, Rob High, a spus unui reporter de la Techrepublic: „Cu orice tehnologie nouă, este foarte important să ne gândim cum o vom crea într-un mod etic şi responsabil. Pentru noi, totul se rezumă la trei principii de bază: încredere, respect şi spaţiu privat... Bineînţeles, spaţiul privat presupune că recunoşti că datele tale sunt datele noastre”[90].

Nu aş fi putut inventa aşa ceva! „Datele tale sunt datele noastre”? Nu, nu este aşa. Datele tale nu sunt datele IBM-ului. Poate că i-a scăpat pe moment, dar cuvintele lui High exprimă minunat sentimentul că li se cuvine orice pe care îl au companiile din domeniul tehnologiei. Cel mai probabil va fi nevoie de reglementări, dar până în anul 2062, datele tale trebuie să fie recunoscute universal ca fiind datele tale.

IA ca soluţie

Ca în multe alte domenii, IA nu este doar o parte din problemă, ci şi o potenţială parte semnificativă din soluţie. Există un număr de modalităţi prin care o IA poate ajuta la prezervarea vieţii private. Una dintre cele

[90] Jason Hiner, „IBM Watson CTO: The 3 Ethical Principles AI Needs to Embrace“, *TechRepublic*, 2 martie 2018.

mai sigure modalităţi este să nu permiţi ca datele să plece din posesia ta.

Până în anul 2062, vom avea suficientă putere computaţională pe dispozitivele noastre, iar computaţia va putea avea loc chiar acolo. Telefonul tău mobil va fi suficient de inteligent încât să-ţi poată recunoaşte vocea, să înţeleagă ce îi ceri şi să acţioneze, fără să apeleze la Google sau la oricare alt serviciu din cloud. Monitorul tău de sănătate nu va trebui să transmită date despre funcţiile tale vitale lui Fitbit şi nimănui altcuiva. Îţi va monitoriza pulsul şi va identifica de unul singur când vei avea nevoie să mergi la doctor. Până în 2062 vom avea asistenţi IA pentru viaţă privată şi securitate pe toate dispozitivele pe care le folosim. Unicul lor scop va fi să ne protejeze intimitatea şi să ne apere securitatea. Ne vor monitoriza toate datele de intrare şi de ieşire, şi vor interveni atunci când intimitatea sau securitatea noastră va fi în pericol.

Alte tehnologii vor contribui de asemenea la păzirea intimităţii noastre. De exemplu, criptografia cuantică va fi un lucru obişnuit şi va oferi o protecţie şi mai mare asupra datelor noastre. Iar tehnologii precum intimitatea diferenţială se vor maturiza şi ne vor permite să schimbăm date cu alţi membri ai societăţii pentru binele public, dar fără să renunţăm la

intimitatea noastră. *Homo digitalis* ar putea să aibă mult mai multă intimitate faţă de *Homo sapiens.* Dacă vom lua deciziile corecte, intimitatea nu va fi o anomalie istorică. Va fi un drept oferit de tehnologie.

0009
Sfârşitul politicii

Un domeniu în care spaţiul privat este foarte important este cel al politicii. Avem nevoie de spaţii private în care să putem explora alternativele la statu-quoul din prezent. Dar chiar dacă reuşim să folosim cu succes IA şi alte tehnologii noi pentru a ne proteja intimitatea, politica va arăta complet diferit în anul 2062. Şi în funcţie de deciziile noastre, s-ar putea să nu fie diferit în sensul bun al cuvântului.

De-a lungul ultimului deceniu am văzut un număr de exemple de tehnologii, în general, şi de reţele de socializare, în special, care au început să modifice dezbaterile politice. La început, a părut ceva pozitiv. Internetul ne ajută să ne conectăm cu alţi oameni în diferite feluri şi a oferit o voce multor persoane care nu au mai avut această şansă înainte.

În anul 2011 am avut parte de o avanpremieră a acestui potenţial pozitiv, când o pagină anonimă de Facebook, intitulată „Cu toţii suntem Khaled Saeed”, a

contribuit la declanşarea revoluţiei egiptene. Khaled Mohamed Saeed a fost un tânăr egiptean ucis în timp ce se afla în custodia poliţiei în Alexandria, în iunie 2010. Pagina de Facebook care vorbea despre această moarte violentă a devenit virală şi în scurt timp a strâns 100.000 de urmăritori. Pagina i-a determinat pe egipteni să protesteze pe 25 ianuarie, ziua sărbătorii naţionale a poliţiei egiptene. Zeci de mii de oameni au ieşit în stradă. După încă 17 zile de demonstraţii, cu sute de mii de protestatari care mărşăluiau în Cairo şi în alte oraşe egiptene, vicepreşedintele Omar Suleiman a anunţat că Hosni Mubarak îşi va da demisia din funcţia de preşedinte.

Au fost şi alte moduri în care reţelele de socializare au jucat un rol important în această revoluţie. Unul dintre protestatari, Fawaz Rashed, a scris pe Twitter:

Folosim Facebook ca să organizăm protestele, Twitter ca să ne coordonăm şi Youtube ca să dăm de veste lumii.
#egipt #25ian.

Mulţi încep să vadă reţelele de socializare ca pe o forţă puternică şi pozitivă ce va duce la schimbări politice. Nu doar cei care au putere pot să vorbească la scară largă: oricine are conexiune la internet poate să

aibă acces la un public numeros.

Tehnologia şi politica

Bineînţeles, tehnologiile noi de comunicaţii au fost cooptate adesea în scopuri politice. Începând cu secolul al şaisprezecelea, tiparul a fost folosit pentru producţia de pamflete politice. Pamfletul lui Thomas Paine din anul 1776, *Common Sense,* a prezentat argumente în favoarea independenţei americane, ajutând la unificarea coloniei în spiritul independenţei. Este considerat a fi cea mai importantă scriere din timpul Revoluţiei Americane.

Mai recent, radioul i-a adus pe politicieni în casele oamenilor. Winston Churchill a scris şi a ţinut numeroase discursuri memorabile în timpul celui de-al Doilea Război Mondial, care au ajutat la obţinerea victoriei. Unul dintre cele mai memorabile a fost discursul pe care l-a ţinut la Camera Comunelor pe data de 4 iunie 1940, ce a urmat după înfrângerea de la Dunkirk. Cuvintele lui Churchill au dat curaj unei naţiuni care se aştepta la o invazie a naziştilor în orice moment: „Trebuie să ne luptăm pe mări şi pe ocean, trebuie să luptăm cu din ce în ce mai multă încredere şi forţă în aer, trebuie să ne apărăm insula, oricare ar fi

costul... nu trebuie să ne predăm niciodată!”[91] Cine nu ar fi mişcat de cadenţa faimoasă a lui Churchill?

Şi este imposibil să îţi imaginezi cum ar arăta politica în ziua de azi fără televiziune. Primele dezbateri prezidenţiale televizate, dintre senatorul John F. Kennedy şi vicepreşedintele Richard Nixon din anul 1960, i-au ştirbit popularitatea lui Nixon şi l-au catapultat pe Kennedy spre postul de preşedinte.

Fake news

Nu este surprinzător că tehnologiile de comunicaţii noi au un impact asupra politicii. Şi din moment ce mulţi oameni îşi preiau ştirile de pe reţelele de socializare, una dintre cele mai mari griji este cea a ştirilor false. După alegerea lui Donald Trump în Statele Unite, Mark Zuckerberg, fondatorul şi CEO-ul Facebook, a respins iniţial ideea cum că ştirile false ar fi jucat vreun rol în treaba asta. „Personal consider că ideea că ştirile false de pe Facebook, care reprezintă o cantitate foarte mică din conţinutul de pe platformă, au influenţat în vreun fel alegerile este o idee destul de

[91] Nu foarte multă lume îşi dă seama de faptul că discursul lui Winston Churchill de pe data de 4 iunie 1940 nu a fost înregistrat. Un crainic a citit extrase din el la ştirile din acea seară de pe BBC. Churchill a produs o înregistrare audio în anul 1949, pe care au auzit-o foarte mulţi şi au crezut că era din anul 1940 de fapt.

trăsnită", a declarat el. „Votanţii iau decizii pe baza experienţei lor de viaţă."[92]

După ce dovezile că lucrurile nu stăteau deloc aşa au început să se adune (iar unii au supranumit site-ul „Fakebook"[93]), Zuckerberg a dat înapoi în februarie 2017 şi a publicat un manifest de 6000 de cuvinte în care a acceptat faptul că Facebook era şi el responsabil. Iar una dintre soluţiile principale pe care le-a propus pentru problema ştirilor false a fost IA. Având în vedere numărul mare de postări, a susţinut el, singura speranţă ca Facebook să filtreze conţinutul la scară globală o reprezintă algoritmii inteligenţi.

Încercările Facebook de până acum în privinţa combaterii ştirilor false prin controlori umani au avut un impact limitat. IA poate ajuta la detectarea ştirilor false aşa cum speră Zuckerberg – dar s-ar putea să şi înrăutăţească problema în acelaşi timp. Algoritmii similari cu cei care detectează ştirile false vor reuşi să genereze ştiri false. Şi pe măsură ce aceşti algoritmi care generează *fake news* vor deveni mai inteligenţi, va fi din ce în ce mai greu să ne dăm seama ce este adevărat şi ce nu. Adevărul va fi victima acestei bătălii în cele din urmă.

---

[92] Casey Newton, „Zuckerberg: The Idea that Fake News on Facebook Influenced the Election Is Crazy", *The Verge*, 10 noiembrie 2016.

[93] Joc de cuvinte în limba engleză între *face,* faţă, şi *fake,* fals. (n. red.).

Facebook nu este singura companie de tehnologie care a fost criticată. Youtube şi Twitter au fost şi ele acuzate de distorsionarea dezbaterilor politice. Dar Facebook a fost conştientă de mulţi ani de puterea pe care o are în ceea ce priveşte alegerile. Acest aspect este extrem de îngrijorător mai ales când despre fondatorul şi CEO-ul său se spune că ar putea deveni preşedintele SUA pe viitor.

Facebook a ştiut

În 2010, cercetătorii de la Facebook şi de la Universitatea California din San Diego au realizat un experiment pe 61 de milioane de membri ai societăţii americane, fără ca ei să ştie acest lucru, în timpul alegerilor locale din SUA. Ştim acest lucru deoarece cercetătorii au publicat rezultatele după doi ani în prestigiosul jurnalul de ştiinţă Nature1. Obiectivul experimentului a fost nobil aparent: creşterea participării la vot.

Experimentul a fost realizat pe toţi oamenii cu vârste peste 18 ani în Statele Unite care au folosit Facebook pe data de 2 noiembrie 2010, ziua alegerilor. Utilizatorii au fost împărţiţi în trei grupuri alese aleatoriu. Primul grup a primit un mesaj care anunţa „Astăzi sunt alegerile”; alţii au văzut acelaşi mesaj alături de câteva

pictograme cu prietenii lor care votaseră şi care aveau scris cuvintele „Am votat”; iar celui de-al treilea grup nu i s-a arătat nimic. Rezultatele cercetătorilor au sugerat că intervenţia lor a sporit prezenţa la vot cu 340.000 de voturi. Acest număr a reprezentat în jur de 0,5% din numărul total de voturi înregistrate. Experimentul făcut de Facebook nu şi-a propus să modifice rezultatul, ci doar să crească prezenţa la vot. Oamenii nu au fost îndemnaţi să voteze pe cineva anume. Utilizatorii din cele trei grupuri diferite au fost aleşi complet aleatoriu. Care ar putea fi problema în cazul ăsta?

Ei bine, să ne gândim la Districtul Windsor-Orange[94] pentru Camera Reprezentanţilor din Vermont. Alegerile din 2010 din acest district au fost decise de un singur vot. Rezultatul alegerilor din Districtul Rutland 5-4 pentru Camera Reprezentanţilor din Vermont a fost şi el decis de un singur vot. Ambele alegeri au fost câştigate de o membră a Partidului Democrat, împotriva unui candidat republican. Într-o astfel de cursă strânsă, experimentul de pe Facebook ar fi putut fi critic.

---

[94] Vezi Robert M. Bond, Christopher J. Fariss, Jason J. Jones, Adam D.I. Kramer, Cameron Marlow, Jaime E. Settle & James H. Fowler, „A 61-million-person Experiment în Social Influence and Political Mobilization”, *Nature*, vol. 489, 2012, pp. 295-298.

Să presupunem pentru o clipă că în 2010 eşantionul experimentului Facebook ar fi fost mai tânăr şi mai feminin în Vermont decât media populaţiei din Vermont. Aceasta nu este o presupunere iraţională: Facebook atrage mai mult tinerele cu vârste cuprinse între 18 şi 29 de ani. Să presupunem că cel mai probabil tinerele din Vermont ar fi votat o femeie din partidul democrat în locul unui bărbat din partidul republican. Din nou, nu este o ipoteză absurdă. Aşadar, creşterea prezenţei la vot a utilizatorilor Facebook din Vermont ar fi putut cu uşurinţă să le aducă un vot sau două în plus democraţilor. Deoarece aceste alegeri au fost atât de contestate, rezultatul ar fi putut fi schimbat în acest mod, transformând o probabilă victorie a republicanilor în victoria democraţilor, ceea ce s-a şi întâmplat.

Cercetătorii care au realizat acest experiment nu ar fi trebuit să fie surprinşi că se putea întâmpla aşa ceva. Mii de alegeri diferite s-au desfăşurat pe 2 noiembrie 2010, dintre care unele au fost cu siguranţă contestate. Într-adevăr, Vermont era unul dintre locurile în care cel mai probabil aveau să existe rezultate apropiate. Camera Reprezentanţilor din Vermont are un număr de electori relativ mic, ceea ce face posibil ca rezultatul să fie unul strâns. În 1977, 1986 şi 2016 au existat alte districte din cadrul Camerei Reprezentanţilor din

Vermont unde rezultatul alegerilor a fost decis tot printr-un singur vot.

Facebook a realizat şi alte experimente pentru creşterea participării la vot la alegerile americane din anul 2012. Nu se ştiu prea multe despre aceste experimente, deoarece nu s-a scris despre ele într-o publicaţie ştiinţifică. Facebook susţine că a ales votanţi aleatoriu şi că nu s-a concentrat pe un anumit grup[95]. Dar încă o dată, Facebook nu reprezintă un eşantion demografic reprezentativ pentru populaţia de votanţi din alegerile americane. Realizarea unui alt experiment pe utilizatori Facebook avea să influenţeze cel mai probabil alegerile din anul 2012[96].

Campanii ţintite

Experimentele celor de la Facebook de a creşte

---

[95] „Întotdeauna am implementat aceste teste (pentru a încuraja participarea la vot) într-o manieră neutră. Şi am învăţat din experienţă proprie şi suntem 100% dedicaţi unei transparenţe şi mai mari ori de câte ori vom încuraja participarea civică pe viitor“, a spus Michael Buckley, vicepreşedintele pentru comunicaţii globale de la Facebook. Citatul acesta a apărut în articolul „Facebook Wants You to Vote on Tuesday. Heres How It Messed With Your Feed in 2012“ din *Mother Jones*, al lui Micah L. Sifry, pe 31 octombrie 2014.

[96] În mod oficial, la alegerile din anul 2012, alegerile din Districtul 27 pentru Camera Reprezentanţilor din statul New Mexico au dus la egalitate. După o renumărare, candidatul republican a fost declarat câştigător la o diferenţă de opt voturi. Este destul de plauzibil ca experimentele Facebook din anul 2012 să fi influenţat unele dintre rezultate, la fel cum s-a întâmplat probabil şi în 2010.

prezenţa la vot din anii 2010 şi 2012 au fost nişte campanii la scară largă. Au vizat milioane de votanţi, aleşi la întâmplare. Poate mai îngrijorător este faptul că reţelele de socializare pot să ţintească grupuri mici de oameni, într-un mod foarte concentrat şi cu puţine costuri. Şi se ştie de ceva timp că Facebook poate să facă asta foarte bine.

În martie 2011, o campanie politică desfăşurată online de Chong + Koster, o agenţie de publicitate digitală, a câştigat un premiu pentru Cea mai bună utilizare a Noilor Tehnologii, din partea Asociaţiei Americane a Consultanţilor Politici. Campania a durat două luni, începând cu septembrie 2010. S-a limitat la două dintre cele mai populate districte din Florida – Dade şi Broward. Aceste două districte aveau împreună o populaţie de 4,2 milioane de oameni. Scopul campaniei Facebook a fost să respingă o măsură introdusă în cadrul unui referendum privind permiterea unor clase mai numeroase în şcolile publice din Florida. Campania a ţintit grupuri pe care măsura respectivă le afecta direct, cum ar fi părinţii şi educatorii. Un sondaj care a avut loc după alegeri a arătat că a fost o diferenţă de 19% între oamenii care au votat în zonele în care Facebook a difuzat reclame şi cei din zonele care nu au fost ţintite de aceste reclame. Oamenii din zonele expuse la reclame au votat cu 17%

mai mult împotriva propunerii decât cei din zonele care nu au fost vizate de reclame.

Campania a fost remarcabil de ieftină. Costul a fost echivalent cu cel al unei campanii poştale care ar fi vizat doar 200.000 de votanţi. Prin comparaţie, reclamele digitale de pe Facebook au obţinut 75 de milioane de expuneri printre oamenii din zone-cheie din Florida, un utilizator din aceste zone văzând în medie reclama respectivă de cinci ori pe zi. Poate că nu este surprinzător că propunerea a fost respinsă. Facebook şi-a dat seama imediat de capacitatea agenţiei digitale de publicitate de a manipula voturi într-o manieră ţintită. În august 2012, pagina oficială Government and Politics a Facebook a menţionat campania într-un mod elogios: „Chong + Koster este de părere că strategia de a folosi Facebook ca instrument de cercetare a pieţei şi ca platformă pentru suprasaturarea cu reclame poate fi utilizată pentru a influenţa opinia publică în cadrul oricărei campanii politice. Agenţia a aplicat deja acest model şi pentru alte campanii”[97]. Ceva mai clar de atât nu se poate. Facebook poate fi folosit pentru influenţarea opiniei

[97] Vezi „Case Study: Reaching Voters with Facebook Ads (Vote No on 8)“, *Facebook*, facebook.com/notes/us-politics-on-facebook/case-study-reaching-voters-with-facebook-ads-vote-no-on-8/10150257619200882

publice în *orice* campanie politică. Aceste declaraţii rămân online pe pagina Facebook Government and Politics.

Acum există companii care au ca model de afaceri, poate controversat, să facă profit din influenţarea alegătorilor, folosind reclame într-un mod foarte ţintit, bazându-se pe date extrase în moduri dubioase de pe reţelele de socializare şi din alte surse. O astfel de companie care a intrat în atenţia publicului la începutul anului 2018 a fost Cambridge Analytica. Îngrijorarea publicului în legătură cu această companie s-a concentrat pe felul în care obţinea informaţii personale ale alegătorilor americani prin chestionare de pe Facebook, fără ca ei să ştie. Datele erau importante, dar nu erau de fapt decât o parte a ecuaţiei.

Companiile precum Cambridge Analytica nu au ţintit un public anume doar pe baza unor cuvinte-cheie, ci au folosit modele sofisticate ale personalităţii votanţilor pentru mesajele lor politice ţintite. Cambridge Analytica a făcut această declaraţie înspăimântătoare:

Cu până la 5000 de date punctuale despre peste 230 de milioane de alegători americani, construim publicul tău ţintă personalizat, după care folosim aceste informaţii cruciale pentru a îi atrage atenţia, a-l

convinge şi a-l motiva să acţioneze... Pornind de la o cunoaştere fără precedent a electoratului tău, vom identifica exact alegătorii care vor întoarce lucrurile în favoarea ta, vom comunica într-un mod creativ cu ei şi îi vom determina să meargă la secţiile de votare.[98]

Dacă faci calculele, Cambridge Analytica are în jur de un trilion de date despre alegătorii americani. Această cantitate enormă de informaţii i-a oferit companiei o capacitate fără precedent de a identifica alegătorii cu precizie. Bineînţeles, toate partidele din cadrul unei campanii politice pot folosi astfel de tehnologii pentru a încerca să influenţeze votul în favoarea lor. Dar costul va fi unul mare: un electorat divizat şi polarizat.

Fondatorul Facebook spune că îşi doreşte să unească lumea[99]. Dar vânzarea reclamelor politice care ţintesc minorităţi le oferă oamenilor puterea să facă exact opusul: să divizeze comunităţile şi să îi îndepărteze pe oameni şi mai mult unii de alţii.

Până acum, Facebook a fost un jucător foarte activ în ceea ce priveşte polarizarea lumii noastre. Thérèse Wong, „guru digital" în cadrul campaniei lui Trump, a

---

98 Vezi Cambridge Analytica, „About Us", ca-politicl.com/ca-advantage.

99 Mark Zuckerberg, „Bringing the World Closer Together", post pe Facebook, 22 iunie 2017, facebook.com/notes/mark-zuckerberg/bringing-the-world-closer-together/10154944663901634

descris într-un film realizat de BBC în anul 2017 cum a ajutat Facebook la găsirea alegătorilor [100]. Ea a menţionat sediul din San Antonio al Cambridge Analytica, unde angajaţii de la Facebook ajutau Cambridge Analytica să strângă alegători. „Când bagi milioane şi milioane de dolari în platformele astea, o să fii tratat cât de bine se poate", a spus ea pentru BBC. „Fără Facebook, nu am fi câştigat. Vreau să spun că Facebook chiar ne-a oferit un avantaj."

Dacă nu suntem atenţi, politica din anul 2062 va fi decisă cu ajutorul unor astfel de tehnologii. Datele numeroase obţinute de pe reţelele de socializare şi din alte locuri vor fi folosite de cei ce posedă cei mai inteligenţi algoritmi, cu care pot obţine putere şi influenţă. Faptul că o persoană ca Donald Trump a ajuns preşedintele Statelor Unite sau că rezultatul Brexitului a provocat atâta discordie ar putea fi doar începutul.

Boţii

Nu doar Facebook duce la transformarea naturii discursului politic. Şi alte site-uri de reţele de socializare au o influenţă considerabilă. Un site care a

[100] Vezi *Secrets of Silicon Valley, Part 2: The Persuasion Machine*, BBC Two, 13 august 2017.

avut un impact masiv este Twitter. Are o putere atât de mare, încât au existat solicitări ca preşedintele Trump să nu mai aibă acces pe platformă[101]. De fapt, nu oamenii, ci calculatoarele au avut unele dintre cele mai mari efecte pe Twitter.

Donald Trump are în jur de 48 de milioane de urmăritori pe Twitter. Dar este estimat că în jur de 14 milioane dintre aceste conturi sunt false, generate de programe-robot sau boţi. *New York Times* stă cam la fel: din cei 41 de milioane de urmăritori, se estimează că 11 milioane sunt falşi. Este amuzant faptul că în cazul papei Francisc este şi mai rău decât în cazul lui Trump şi al ziarului. Din cei 17 milioane de urmăritori ai papei, mai mult de jumătate – aproape 10 milioane – sunt falşi. Aşa cum era de aşteptat, chiar mai rău stă preşedintele Federaţiei Ruse. Din cei 2,5 milioane de urmăritori ai lui Puţin, în jur de 60%, sau 1,5 milioane, sunt falşi[102].

Unul dintre pericolele acestor boţi este că până în anul 2062 va fi greu pentru oameni să fie auziţi peste mulţimea de voci de calculatoare. Într-adevăr, boţii au

---

[101] În noiembrie 2017, contul de Twitter al lui Trump a fost dezactivat de către un angajat răzvrătit care era în ultima sa zi de lucru la companie. Unii au cerut ca angajatul să fie nominalizat la Premiul Nobel pentru Pace.

[102] Este corect din partea mea să raportez faptul că, din cei 1476 de urmăritori de pe contul meu de Twitter (@TobyWalsh), 41 sunt falşi. N-am nicio idee cum de 41 de boţi au ajuns să mă urmărească, şi de ce s-ar deranja să facă asta.

deja un impact. În anul 2017, Comisia Federală de Comunicaţii a Statelor Unite (CFC) a invitat lumea să lase comentarii despre chestiunea controversată a neutralităţii internetului. Asta înseamnă că toate datele de pe internet ar trebui să fie tratate egal. Nu se poate ca e-mail-urile tale să aibă prioritate în faţa filmuleţului meu de pe Youtube sau a transmisiunii conversaţiei de pe Whatsapp a altei persoane. Din cele 22 de milioane de comentarii pe care CFC le-a primit cu privire la neutralitatea de pe internet, în jur de 80% proveneau de la boţi[103]. Oamenii care au postat comentarii erau majoritatea în favoarea neutralităţii; poate că nu este surprinzător că boţii care au postat comentarii erau cei mai mulţi împotriva neutralităţii[104].

Cu toate acestea, există un bot pe care îl aplaud. O firmă de securitate informatică din Noua Zeelandă, Netsafe, a creat un chat bot care se numea Re: scam. Ori de câte ori primeşti un e-mail de la un escroc din Nigeria, trebuie să îl redirecţionezi către me@rescam.org. Atunci, botul va răspunde escrocului în locul tău, şi astfel îl va face să-şi piardă timpul.

Calculatoarele care pretind că sunt oameni au o

---

[103] Jane Wakefield, „Net Neutrality Debate - Controlled by Bots“, BBC News, 4 octombrie 2017.

[104] Din punctul meu de vedere, singurul argument bun împotriva neutralităţii internetului este că datele generate de oameni ar trebui să aibă prioritate în faţa datelor generate de boţi!

istorie lungă în domeniul IA. Într-adevăr, faimosul test al lui Alan Turing prin care se poate dovedi că un calculator gândeşte îi cere să imite un om. În fiecare zi ni se cere să completăm teste CAPTCHA ca să dovedim că suntem într-adevăr oameni.

Dar pe măsură ce IA devine mai performantă, va fi din ce în ce mai greu să ne dăm seama care sunt calculatoarele şi care oamenii. De fapt, deja ne apropiem de acel moment din cauza lui Donald Trump. Bradley Hayes, cercetător la Laboratorul de Cercetare Informatică şi Inteligenţă Artificială de la MIT, s-a folosit de învăţarea mecanică pentru a construi un bot pe Twitter, numit ©Deepdrumpf[105]. L-a antrenat folosind bucăţi din discursurile lui Trump. Acum postează tweeturi care ar putea foarte bine să fie chiar ale preşedintelui.

20 ian 2017

[Vom fi protejaţi de Dumnezeu.] Nu vom avea de câştigat cu serviciile medicale. Nu ni le putem permite.

Este foarte simplu. Obamacare este un dezastru.

#inaugurare

---

[105] Chatul DeepDrumpf este numit după o secvenţă din emisiunea *Last Week Tonight*, în care John Oliver i-a încurajat pe oameni să vorbească despre Trump folosind numele său de familie original, Drumpf.

20 ian 2017

Ca să nu fie nicio confuzie, intenţia mea nu este să mă descotorosesc de guvern. Ci să-l fac să fie groaznic, în mare.

26 sept 2016 Răspuns către @joss

(Minciuni), adică chiar sunt urmăriţi penal? Face cineva ceva? Asta o să mă ajute să ajung în Biroul Oval.

@joss @democraţii #dezbateri #searadezbaterilor

Politicieni falşi

Poate că nu te-a convins @Deepdrumpf? Dar până în anul 2062 nu vei mai putea să îţi dai seama care sunt politicienii adevăraţi şi care cei falşi. De asemenea, nu vei mai putea să îţi dai seama care politicieni umani ţin discursuri adevărate şi care false. Deja ştim că nu putem avea încredere în fotografii. Programe ca Photoshop pot să adauge sau să şteargă oameni cu uşurinţă dintr-o poză. În curând, nu vei mai putea avea încredere în nicio înregistrare audio sau video.

În 2016, Adobe a oferit o demonstraţie cu Voco, un software pentru editare şi generare de sunete. A fost numit „Photoshoppentru-voce”. Pornind de la un discurs real al unei persoane, Voco arată o transcriere unde poţi pur şi simplu să tai sau să adaugi text. Nici că

ar putea fi mai uşor să schimbi cuvintele unui politician. Alte companii, cum ar fi Candyvoice şi Lyrebird, se grăbesc să dezvolte instrumente software similare.

Înregistrările video vor urma o cale asemănătoare. Deja există software pentru a lipi o faţă pe un alt corp într-un film. În curând, vei putea să introduci o persoană întreagă într-o scenă. Şi nu vorbim doar despre software care să manipuleze filmuleţe existente; într-o zi vom avea programe care pot crea scene complet artificiale imposibil de deosebit de cele reale.

Până în anul 2062, nu vei mai putea să crezi în nimic din ce vezi sau ce auzi decât dacă eşti chiar acolo. Din păcate, politicienii lipsiţi de scrupule vor profita de pe urma acestor instrumente noi. Pur şi simplu vor nega veridicitatea oricărei înregistrări audio sau video care iese la iveală şi care îi face de râs. Se pare că adevărul va deveni un concept fungibil.

A patra putere

Una dintre cele mai importante responsabilităţi ale presei este să dezvăluie adevărul. Să expună corupţia şi minciunile. Să se asigure că politicienii rămân sinceri. În anul 2062, mass-media îşi va îndeplini cu greu această misiune fundamentală şi esenţială. Putem

spune deja că lupta împotriva ştirilor false a fost pierdută.

Internetul reprezintă o lovitură triplă asupra practicii jurnalismului. Prima lovitură a fost pierderea venitului obţinut în urma reclamelor. Ziarele plăteau pentru jurnalism în parte prin venitul generat datorită anunţurilor şi reclamelor. Dar companii precum Google şi Facebook au furat o mare parte din acel venit. În mod asemănător, radioul şi televiziunea au pierdut o mare parte din venitul lor obţinut din reclame, în favoarea internetului. În anul 2017, cheltuielile pe reclamele digitale din Statele Unite depăşeau cheltuielile pe reclamele de la televizor pentru prima dată.

A doua lovitură a fost o reducere a venitului obţinut direct de la consumatorii de conţinut. Ziarele au pierdut bani deoarece publicul a început să se aştepte ca tot conţinutul să fie livrat pe gratis online. Mulţi dintre noi, cei care înainte cumpăram ziarul zilnic, nu mai facem asta acum. Şi în timp ce unele ziare cum ar fi *New York Times* au înlocuit această sursă de venituri cu abonamente digitale, mulţi nu au făcut asta.

În Statele Unite, tirajul cotidienelor aproape că s-a înjumătăţit, de la peste 63 de milioane în 1970 la sub 35 de milioane în anul 2016. Astfel de declinuri s-au observat şi în alte ţări. În Regatul Unit, numărul

cititorilor ziarului *Daily Mirror* s-a înjumătăţit în ultimul deceniu, în timp ce *Guardian* şi *Daily Telegraph* au pierdut în jur de 25% din cumpărători. În acelaşi timp, posturile de televiziune au pierdut o parte mare din public în favoarea serviciilor online de difuzare de filme. De pildă, audienţa televiziunii din Statele Unite în rândul persoanelor cu vârste cuprinse între 18 şi 24 ani aproape că s-a înjumătăţit în ultimii cinci ani. Aceşti clienţi s-au mutat de la televiziune la serviciile digitale.

A treia lovitură a fost reducerea numărului de jurnalişti, întrucât companiile media au fost nevoite să scadă cheltuielile pentru a face faţă declinului de venituri. Algoritmii IA sunt folosiţi din ce în ce mai mult pentru a înlocui jurnalişti umani. Şi deşi aceşti algoritmi s-ar putea descurca foarte bine să scrie un articol scurt, nu sunt capabili să producă reportaje lungi şi nici investigaţii.

Există o rază de speranţă că o parte a veniturilor ar putea reveni. Giganţii internetului primesc din ce în ce mai multe solicitări să plătească pentru utilizarea conţinutului de presă pe platforma lor. Şi instituţii cum ar fi Uniunea Europeană se gândesc cum să determine companiile de tehnologie să plătească pentru milioanele de articole de ştiri şi de linkuri care apar pe site-urile lor. Aceste probleme trebuie să fie rezolvate cu mult înainte de 2062 dacă vrem să avem o a patra

putere funcţională, care să-i tragă la răspundere pe politicieni.

Zuckerberg ca preşedinte

Până în anul 2062, liderii companiilor de tehnologie vor fi personalităţi politice importante. În 2017, Mark Zuckerberg a anunţat că obiectivul lui pentru noul an era să „viziteze şi să întâlnească oameni din toate statele americane". Apoi l-a angajat pe strategul-şef al lui Hillary Clinton din timpul campaniei ei prezidenţiale din anul 2016. De asemenea, a modificat statutul companiei sale, astfel încât Facebook să îi permită să deţină o funcţie politică. Nu e de mirare că multe persoane s-au întrebat dacă într-o zi americanii l-ar putea avea pe Zuckerberg drept preşedinte.

Mark Zuckerberg s-a născut în White Plains, New York, aşa că are tot dreptul, la fel ca oricare alt cetăţean născut în Statele Unite, să candideze pentru preşedinţie. Dar mulţi şi-au arătat îngrijorarea legată de faptul că o persoană ca Zuckerberg şi-ar dori să obţină un post atât de important. Accesul nerestricţionat la o platformă precum Facebook oferă un avantaj imens unui candidat la alegeri. Ne dorim să avem o lume în care nu câştigă politicienii care au cele mai bune idei, ci cei care au acces la cele mai multe date

şi cei mai buni algoritmi? Multe ţări au limite în ceea ce poate fi donat candidaţilor şi ce poate fi cheltuit de aceştia. Astfel de reguli previn ca alegerile să fie câştigate pur şi simplu de cei care au cei mai mulţi bani. Până în anul 2062 vom avea nevoie de legi similare care să limiteze folosirea datelor şi a algoritmilor în timpul alegerilor.

S-ar putea să fie necesare şi legi care să limiteze influenţa *lobbyingului.* În ultimii ani, companiile de tehnologie au trecut de la o prezenţă nulă la Washington la statutul de cei mai mari lobbyişti. Potrivit dosarelor federale, în anul 2017 Google a cheltuit mai mult pe lobbying – 18 milioane de dolari – decât oricare altă companie. Facebook a fost puţin în urma sa, cu 11,5 milioane de dolari, iar Amazon cu 12,8 milioane, Microsoft cu 8,5 şi Apple cu 7 milioane de dolari. Cu excepţia lui Microsoft, fiecare dintre aceste companii a cheltuit cu 2 sau 3 milioane de dolari mai mult pe lobbying decât în anul precedent, 2016.

Alegeri politice

Criticul literar Roland Barthes a observat că tehnologia este mitică. Tindem să ne gândim la tehnologie ca şi cum ar fi un dar de la Dumnezeu, o parte a ordinii fireşti a universului. Uităm de faptul că

este produsul unui anumit context politic şi istoric. Tehnologia nu este inevitabilă: este ceea ce alegem noi să fie.

Să luăm ca exemplu televiziunea. Când a fost fondat BBC, guvernul britanic a decis că va fi un bun public, finanţat de taxe şi nu plătit (şi aşadar condus) de interese comerciale. Aceasta a fost o alegere. Consider că a fost o alegere bună. Am fi putut să decidem cu 70 de ani în urmă şi că televiziunea este mult prea frivolă pentru politică. Imaginaţi-vă cât de bune ar fi dezbaterile politice de acum dacă nu ar avea loc prin sloganuri la televizor. Dacă politica ar mai fi încă discutată prin articole serioase în ziare şi prin adunări publice.

Până în 2062, s-ar putea să decidem că reclamele politice ar trebui să fie interzise pe reţelele de socializare. Sau că partidelor politice le va fi permis să-şi difuzeze mesajele către publicul larg, dar nu să le disemineze unui public restrâns. Sau că boţii politici nu vor mai avea acces pe Twitter – sau vor fi interzişi complet pe internet. Într-un discurs excelent pe care l-a ţinut în anul 1998, umanistul Neil Postman a declarat:

Cel mai bine este să vedem tehnologia ca pe un intrus ciudat, trebuie să ne aducem aminte că tehnologia nu face parte din planul lui Dumnezeu, ci este un produs

al creativităţii şi al aroganţei oamenilor, iar capacitatea ei de a face lucruri bune sau rele depinde în întregime de cât de conştienţi sunt oamenii de ceea ce face tehnologia pentru noi şi cu noi.[106]

Postman a avut dreptate. Tehnologia este un intrus ciudat pe care nu trebuie să-l invităm în toate aspectele vieţii noastre. Trebuie să luăm nişte decizii dificile legate de felul în care IA este folosită în politică, astfel încât până în anul 2062 să îmbunătăţească, şi nu să dăuneze dezbaterile politice. Sper că vom face alegerile corecte, din moment ce deja devenim conştienţi de marele său potenţial pentru a fi folosită greşit.

0010
Sfârşitul occidentului

La începutul anului 2018, primele patru companii din lume după capitalizarea de piaţă au fost toate companii de tehnologie: Apple, Alphabet (din care face parte Google), Microsoft şi Amazon. Software-ul a început cu adevărat să înghită lumea.[107] Google răspunde la opt

---

[106] „Five Things We Need to Know About Technological Changes“ a fost un discurs al lui Neil Postman la Conferinţa New Tech din anul 1998 din Denver, Colorado, pe data de 27 martie 1998.

[107] Vezi Marc Andreessen, „Why Software Is Eating The World”, *The Wall Street Journal*, 20 august 2011. Andreessen este co-fondator al firmei de capital

din nouă căutări la nivel global. Acest rezultat ar fi şi mai mare dacă nu ar fi practic interzis în China. Ceilalţi giganţi de IT sunt dominanţi în propriile nişe. Lunar, unul din patru oameni foloseşte Facebook. Microsoft oferă un sistem de operare pentru mai mult de 80% din laptopurile şi calculatoarele de pe planetă. Iar Amazon este responsabil pentru aproape jumătate din comerţul electronic din America.

Aşadar, poate că veţi crede că aceste patru companii vor domina viaţa în anul 2062, exact cum ne domină vieţile online în prezent. Dar giganţii de IT de azi încep să se confrunte cu o competiţie serioasă. Şi nu din partea unor companii proaspăt înfiinţate extraordinare, care folosesc IA. Este prea uşor pentru aceşti titani să cumpere sau să saboteze orice companie la început de drum. Nu, competiţia vine din China. Companii de tehnologie precum Alibaba, Baidu şi Tencent s-au născut pe piaţa chineză protejată, dar acum se dezvoltă rapid.

În anul 2017, Alibaba Group a anunţat că plănuieşte să investească mai mult de 5 miliarde de dolari în următorii 3 ani în cercetare şi dezvoltare în domeniile IA, Internet of Things şi computaţia cuantică. În anul

---

Andreessen-Horowitz, care a investit în Facebook, Groupon, Skype şi Twitter. Înainte să devină investitor de capital cu risc, a fost unul dintre fondatorii Netscape.

2018, Baidu a solicitat o OPI[108] pentru serviciul său asemănător cu Netflix, despre care se estimează că va strânge 10 miliarde de dolari. Profiturile obţinute în urma vânzării vor fi folosite pentru cercetarea şi dezvoltarea IA. Iar în 2016 Tencent a construit un laborator IA în Shenzen cu aproape 400 de angajaţi, ce face parte din strategia companiei cu privire la IA.

Chiar şi fără aceste investiţii în IA, giganţii de tehnologie din China sunt semnificativi ca mărime. Alibaba este cel mai mare magazin online din China. În prezent are o valoare de 500 de miliarde de dolari. Prin comparaţie, capitalizarea de piaţă a Amazonului este în jur de 750 de miliarde de dolari. Dar Alibaba creşte mai rapid decât Amazon. Între anii 2012 şi 2016, vânzările Amazonului s-au dublat, dar cele ale Alibaba sunt de mai bine de trei ori mai mari. Aşadar, ne putem aştepta ca Alibaba să îşi eclipseze rivalul american. Cât despre Baidu, este cel mai mare motor de căutare din China şi al patrulea cel mai vizitat website din lume. Iar Tencent deţine cea mai utilizată platformă socială din China, Wechat. Compania este estimată la circa 500 de miliarde de dolari, aproape tot atât cât valorează Facebook.

---

[108] Ofertă publică iniţială (n. tr.).

Momente Sputnik

Pe data de 4 octombrie 1957, Uniunea Sovietică a lansat Sputnik, primul satelit artificial care a ajuns în orbita joasă a Pământului. Nu arăta prea impresionant: o sferă simplă şi poleită, cu un diametru de vreo 60 de centimetri. Şi nu putea să facă prea multe, decât să difuzeze transmisii radio simple cu ajutorul celor patru antene externe. Dar a făcut ca Statele Unite să se trezească şi să-şi dea seama de ameninţarea priceperii tehnice a Rusiei, aşa că au pornit spre Lună.

În cursa de a construi IA, cele două victorii ale programului Deepmind la jocul Go împotriva jucătorilor umani par să fie nişte „momente Sputnik". În acest caz, nu Statele Unite au fost cele care s-au trezit, ci Orientul. Şi nu cursa către Lună a fost cea declanşată, ci cursa pentru a construi IA.

La scurt timp după ce sud-coreeanul Lee Sedol a fost bătut de Alphago în martie 2016, guvernul naţiunii sale a creat un fond de un trilion de woni (863 de miliarde de dolari) pentru cercetare în domeniul IA pentru următorii cinci ani. Dar poate că şi mai importantă a fost trezirea gigantului chinez, cu un an mai târziu. După victoria lui Alphago împotriva chinezului Ke Jie, guvernul chinezesc a anunţat un plan ambiţios prin care să devină prima ţară din lume în domeniul IA.

Planul Chinei estimează că, până în anul 2030, IA va contribui cu un trilion de yeni (150,8 miliarde de dolari) direct la producţia industrială, şi cu 20 trilioane de yeni (1,5 trilioane de dolari) indirect prin industriile conexe.

Victoriile lui Alphago au fost suficient de „supărătoare", încât autorităţile chineze au interzis difuzarea în direct a jocurilor pe internet. Go are o importanţă specială pentru China, a fost inventat cu peste 2000 de ani în urmă. Este considerat a fi una dintre cele patru arte esenţiale pe care aristocraţii trebuie să le stăpânească, alături de guqin (un instrument cu corzi), she (caligrafie) şi hua (pictură chineză). Nu e de mirare că i-a şocat foarte tare faptul că pentru prima oară calculatoarele au putut să bată cei mai buni jucători umani de Go.

Planul Chinei

În iulie 2017, la două luni după înfrângerea lui Ke Jie, Consiliul de Stat al Chinei a iniţiat „Planul de Dezvoltare a Noii Generaţii IA". Planul nu prea ascunde ambiţia ţării de a folosi IA pentru a domina economic şi militar lumea întreagă. Aşa cum a declarat preşedintele Xi Jinping în timpul celei de-a nouăsprezecea ediţii a Congresului din octombrie 2017, China are ca scop să devină „superputerea ştiinţei şi a

tehnologiei". Dar planul Consiliului de Stat nu urmăreşte doar puterea economică şi militară; descrie şi ambiţia de a folosi IA pentru a-i controla chiar pe cetăţenii chinezi. IA „va îmbunătăţi semnificativ capacitatea şi nivelul guvernării sociale, jucând un rol de neînlocuit în menţinerea stabilităţii societăţii într-un mod eficient".

Alte părţi ale planului sunt mai puţin problematice. Într-adevăr, unele aspecte – cum ar fi apelul său la protejarea inteligentă a mediului şi la sistemele inteligente pentru protecţia cetăţenilor – sunt demne de laudă. Totodată, este îmbucurător să vedem obiective precum cele de „a sprijini cercetarea în problemele etice, legale şi sociale ce ţin de IA şi a crea legi, reglementări şi standarde etice de lucru pentru a asigura dezvoltarea sănătoasă a IA".

Primul obiectiv concret al Consiliului de Stat a fost ca, până în anul 2020, China să fie competitivă la nivel global în domeniul IA. Aşa cum voi prezenta pe scurt, există un număr de aspecte care sugerează faptul că avea de gând să-şi atingă obiectivul chiar mai devreme. Planul precizează de asemenea ambiţia Chinei de a face descoperiri majore în teoriile de bază ale IA până în anul 2025 şi de a fi prima în domeniul IA până în anul 2030. Sunt destul de sigur că chinezii vor reuşi aceste lucruri. Chiar dacă nu vor respecta aceste termene, cu

siguranţă îşi vor atinge obiectivele până în 2062. Până atunci China va fi cel mai probabil prima putere economică şi militară din lume.

În ultimele decenii, sistemul neobişnuit din China, care combină totalitarismul cu un capitalism controlat a dus la o creştere economică mai rapidă decât a oricărei ţări din Occident. PIB-ul Chinei s-a dublat între 1998 şi 2004, în şase ani, s-a triplat în următorii şase ani, şi aproape că s-a dublat din nou în următorii şase ani. Prin comparaţie, PIB-ul Statelor Unite a crescut cu o treime în fiecare dintre aceste perioade de şase ani. De-a lungul a 18 ani, între 1998 şi 2016, PIB-ul Statelor Unite doar s-a dublat, în timp ce al Chinei a crescut de mai mult de zece ori.

China este fără îndoială o piaţă vastă. Are o populaţie de aproape 1, 4 miliarde de oameni, de patru ori mai mult faţă de cea din SUA. Acest lucru face posibilă scăderea costurilor prin producţia la scară largă în feluri în care America şi Europa nu ar putea visa vreodată. China este clar cea mai mare piaţă a telefoanelor inteligente din lume, cu în jur de 750 de milioane de aparate. Acest număr este triplu faţă de cel din Statele Unite şi dublu faţă de Europa.

China adoptă tehnologia rapid. În 2016, volumul plăţilor prin telefon din China s-a dublat, şi a ajuns la 5 trilioane de dolari. Se estimează că plăţile electronice

din China se vor înmulţi de nouă ori, ajungând la 45 de trilioane de dolari până în 2021. În prezent deja banii lichizi nu sunt acceptaţi în anumite locuri din China. Ca să punem lucrurile în perspectivă, Statele Unite au avut doar 112 miliarde de dolari plătiţi electronic în anul 2016. De 45 de ori mai puţin decât China.

China are alte avantaje în cursa pentru construirea IA. De exemplu, are o atitudine mai „relaxată" faţă de viaţa privată. Companiile de tehnologie mari şi statul chinez pot să facă schimb de informaţii mult mai uşor decât în Occident. Aşa cum am văzut, sistemul social de credite este un bun exemplu. Şi chiar dacă acest lucru este îngrijorător din perspectiva drepturilor omului, colectarea datelor în acest mod nu va face decât să mărească avantajul Chinei în cursa pentru IA.

Dominaţie ştiinţifică

În anul 1999 guvernul chinez a început un program prin care să crească masiv procentul celor cu educaţie superioară. Numai în acel an, numărul celor care au intrat la facultate aproape că s-a dublat. În următorii 15 ani, efectivul studenţilor a crescut rapid, depăşind numărul din Statele Unite din anul 2001; în prezent, China are de două ori mai mulţi studenţi decât SUA. În 2016 China construia aproape o universitate pe

săptămână.

Mulţi dintre aceşti studenţi chinezi învaţă materii ŞTIM. Forumul Economic Mondial a raportat că în 2016 China a avut 4,7 milioane de absolvenţi ŞTIM. În Statele Unite au fost doar 568.000. Până în anul 2030, China va avea mai mult de 40% din toţi absolvenţii ŞTIM la nivel mondial, în comparaţie cu Europa care va avea doar 8%, iar Statele Unite numai 4%.

Pe lângă faptul că a pregătit mai mulţi oameni în domeniile ŞTIM, China a crescut foarte mult şi fondurile destinate cercetării ştiinţifice. În ultimii 15 ani fondurile pentru cercetare din China au crescut de mai mult de şase ori, în timp ce în SUA şi Europa au crescut cu doar 50%. China este responsabilă acum de 20% din cheltuielile pentru cercetări şi dezvoltare la nivel mondial. Se estimează că va depăşi Statele Unite în anul 2020, devenind ţara cu cele mai mari investiţii în cercetare.

Recent, atât guvernul chinez, cât şi afacerile chineze au început să investească foarte mult în cercetarea IA. Din punct de vedere istoric, China nu a fost o forţă majoră în acest domeniu. Cu doar un deceniu în urmă nu se desfăşurau prea multe cercetări pe IA în această ţară. În anul 2013, m-am aflat în consiliul conferinţei internaţionale de IA atunci când s-a decis ca aceasta să aibă loc în China pentru prima dată. Obiectivele

noastre erau simple: să sprijinim cercetarea în domeniul IA în China şi să ne folosim de potenţialul imens pe care îl are această ţară. Deja putem vedea cum este actualizat acel potenţial. După doar patru ani, cercetătorii chinezi au venit cu mai multe proiecte la conferinţa din 2017 decât oamenii de ştiinţă americani şi europeni la un loc. De fapt, cercetătorii chinezi au scris peste o treime din toate rapoartele de cercetare la nivel mondial.

Există şi alte câteva măsuri care sugerează că China a prins din urmă Occidentul, înainte de termenul său limită, anul 2020. CB Insights a raportat că China a avut de şase ori mai multe solicitări de brevete care foloseau cuvintele-cheie „învăţare profundă" decât SUA, şi de cinci ori mai multe care foloseau cuvintele-cheie „inteligenţă artificială"[109]. În plus, companiile proaspăt înfiinţate din domeniul IA din China au obţinut 48% din dolarii investiţi în toate companiile de IA nou apărute din toată lumea în anul 2017, depăşind Statele Unite pentru prima dată.

Răspunsul Americii

Statele Unite au dominat cercetarea din domeniul IA

[109] CB Insights, *Artificial Intelligence Trends to Watch în 2018*, 22 februarie 2018.

încă de la începuturile acestuia. Chiar şi termenul de „Inteligenţă Artificială” a fost folosit pentru prima dată în Statele Unite, în cadrul unui atelier de vară ce a avut loc la Colegiul Dartmouth, în anul 1956. Atunci s-au întâlnit pentru prima dată mulţi dintre cercetătorii care au înfiinţat acest domeniu şi au început să se preocupe cu provocarea construirii de maşini care gândesc.

Deci nu ar trebui să ne surprindă faptul că Statele Unite au un plan prin care să contraatace ameninţarea Chinei de a câştiga cursa în IA[110]. De fapt, SUA are un plan foarte credibil care a fost făcut public în octombrie 2016, cu aproape un an înainte ca planul Chinei să fie dezvăluit. Şi este uşor de observat faptul că multe dintre obiectivele planului american se reflectă în planul chinez.

Planul Americii a fost conceput de *Office of Science & Technology Policy* (OSTP), departamentul din cadrul guvernului Statelor Unite care îl consiliază pe preşedinte cu privire la efectele ştiinţei şi ale tehnologiei asupra afacerilor interne şi externe. Raportul a fost constituit dintr-un proces de consultare complexă a guvernului, universităţilor, industriei şi a publicului.

Planul SUA include 23 de recomandări. Acestea

---

[110] Pregătirea pentru Viitorul Inteligenţei Artificiale, Consiliul Naţional de Ştiinţă şi Tehnologie, octombrie 2016.

variază de la „deschiderea” guvernului şi a datelor sale pentru a fi utilizate de IA la acordarea de prioritate cercetării de bază şi pe termen lung în domeniul IA. Cuprinde şi obiective concrete, cum ar fi cererea unui sistem de control al traficului aerian care să permită gestionarea atât a aparatelor de zbor cu pilot uman, cât şi a celor autonome. Planul cere să existe transparenţă şi corectitudine în utilizarea instrumentelor care folosesc IA, atât din partea guvernului, cât şi a domeniului privat. În cele din urmă, solicită dezvoltarea unor politici unitare, la nivel de guvern, care să ţină cont de dreptul internaţional umanitar, în legătură cu armele autonome şi semiautonome.

Din păcate, recomandările acestui plan au fost ignorate total. La o lună după publicarea raportului, Donald Trump a fost ales preşedinte. Echipa de la OSTP s-a redus cu două treimi, iar cei care au mai rămas nu au cunoştinţele ştiinţifice necesare. Spre deosebire de predecesorul său, se pare că preşedintele Trump nu este prea interesat de politicile din domeniul ştiinţei şi tehnologiei. Directorul de la OSTP a fost în mod tradiţional consilierul preşedintelui în domeniul ştiinţei, dar acest post a rămas neocupat. Ba chiar nici măcar nu a fost nominalizată vreo persoană. Multe alte posturi ce ţin de domeniul ştiinţei şi tehnologiei – inclusiv postul de Ofiţer-Şef în Tehnologie al SUA şi

funcţia de director al DARPA – au rămas şi ele libere. Preşedintele Trump pare că s-a lipsit de statutul Americii de lider în domeniul ştiinţei.

Planurile din alte locuri

Alte ţări au reacţionat de asemenea la planul din domeniul IA al Chinei. Multă lume consideră că Regatul Unit este locul în care s-a născut IA ca domeniu de studiu. Bineînţeles, matematicianul britanic Alan Turing a fost cel care a scris primele articole ştiinţifice despre construirea maşinăriilor inteligente[111]. Iar Marea Britanie continuă şi în ziua de azi să fie un jucător important în cercetarea din domeniul IA. Programul Deepmind de la Google a fost fondat şi continuă să opereze în Londra, spre exemplu.

În noiembrie 2017, profesorul Dame Wendy Hall şi Jérôme Pesenti au publicat o recenzie aprobată de secretarii pentru Afaceri şi Cultură din cadrul guvernului Regatului Unit, legată de cum poate fi dezvoltată IA în ţara lor. Au creat 18 recomandări pentru îmbunătăţirea cantităţii de date şi a abilităţilor, pentru încurajarea cercetării în domeniul IA şi pentru susţinerea folosirii IA. Guvernul britanic a susţinut

[111] Vezi Alan Turing, „Computing Machinery and Intelligence“, *Mind,* vol. 59, nr. 236, 1950, pp. 433-460.

raportul cu promisiunea că în aprilie 2018 va investi 1,3 miliarde de dolari în cercetarea şi dezvoltarea IA. Cu provocările pe care le-a adus Brexitul, sper că aceste iniţiative vor fi de ajuns pentru ca Marea Britanie să rămână competitivă în continuare.

Se preconizează că India va avea o populaţie mai mare decât China până în anul 2022, aşa că merită să ne gândim dacă ar putea reprezenta o ameninţare pentru dominaţia economică în creştere a Chinei. La începutul anului 2018 ministrul de finanţe al Indiei a anunţat în discursul său despre buget că cel mai important grup de experţi în politici publice al guvernului va iniţia un program naţional de IA. Guvernul indian a dublat bugetul alocat în anii 2018-2019 programului India Digitală, acesta ajungând la 30,7 miliarde de rupii (408 milioane de dolari) pentru a susţine această iniţiativă. Deşi este bine că India se menţine în cursă, nici măcar această sumă de aproape o jumătate de miliard de dolari nu va fi de ajuns pentru a câştiga. În anul 2017, oraşul chinez Tianjin, care este al patrulea oraş din China ca populaţie, a anunţat fonduri de 5 miliarde de dolari pentru sprijinirea industriei IA. Chiar dacă aceşti bani vor fi cheltuiţi într-un deceniu, tot vor depăşi cu mult banii cheltuiţi de guvernul indian.

Mai sunt şi alte ţări care au dezvoltat sau sunt în curs de dezvoltare a unor planuri în domeniul IA. Canada a

investit 125 de milioane de dolari în cercetarea din acest domeniu. Chiar dacă au mulţi oameni de ştiinţă excepţionali, tot s-ar putea să nu fie suficienţi bani pentru ca statul canadian să rămână în cursă. Franţa are un plan şi mai agresiv de 1, 85 miliarde de dolari. Iar Uniunea Europeană se ocupă de propriul plan de a menţine Europa în cursă. Cu toate acestea, având în vedere că Brexitul e pe aproape, Europa s-ar putea să piardă cea mai mare şi cea mai puternică comunitate din acest domeniu. Această situaţie nu va fi nici în folosul Europei, nici al Angliei.

Într-un final, am ajuns şi la Australia, care nu a publicat până acum niciun plan privind IA. Este mare păcat, deoarece Australia a avut dintotdeauna o comunitate academică ce a ţintit din ce în ce mai sus în domeniul IA. Australia are şi o comunitate puternică a companiilor proaspăt înfiinţate, ceea ce înseamnă că IA este utilizată în domenii precum robotica, serviciile financiare, medicina şi agricultura. Sper că Australia va rezolva în viitorul apropiat lipsa unui plan[112].

[112] Am acceptat de curând oferta de a ajuta guvernul Emiratelor Arabe Unite să dezvolte un plan privind IA. Motivul pentru care am făcut asta a fost ca să fac de ruşine guvernul australian: cercetătorii australieni dau sfaturi altor ţări legat de ce trebuie să facă, dar Australia nici măcar nu are un plan.

Sfârşitul neoliberalismului

Am afirmat că cel mai probabil China va câştiga cursa IA, în mare parte deoarece Statele Unite şi alte ţări par că pierd. În următorul deceniu, companiile de tehnologie precum Google şi Facebook vor întâmpina din ce în ce mai multă rezistenţă. Acest lucru va stopa creşterea din Occident şi îi va permite Chinei să avanseze din ce în ce mai mult.

Începând cu anii 1980, cu Ronald Reagan în Statele Unite şi Margaret Thatcher în Marea Britanie, ideile neoliberale cum ar fi privatizarea, austeritatea, liberalizarea, comerţul liber şi reducerea din cheltuielile guvernului domină deja de câteva decenii. Dar în vreme ce aceste politici au produs creştere economică atât de mult timp, a existat şi un preţ: creşterea nesiguranţei şi a inegalităţii economice.

Theresa May a susţinut într-un discurs ţinut la Banca Angliei în septembrie 2017 că „o economie de piaţă liberă, care operează după anumite reguli şi reglementări, este cel mai mare agent de progres uman colectiv care a fost creat vreodată". Declaraţia ei implică o condiţie foarte importantă: piaţa trebuie să opereze după regulile şi reglementările corecte. Fără acestea, monopolurile vor distorsiona preţurile. Fără acestea, pieţele nu vor mai plăti externalităţile, cum ar

fi costurile reale de mediu. Fără regulile şi reglementările corecte, pieţele se pot supraîncălzi, ceea ce poate duce la bule speculative. Şi fără regulile şi reglementările corecte, cei care au mai multe informaţii vor avea avantaje pe nedrept.

Prin limitarea regulilor şi a reglementărilor, neoliberalismul expune aceste probleme fundamentale din cadrul economiei de piaţă. Şi unele dintre cele mai nereglementate pieţe care operează azi sunt cele care ţin de tehnologie. Prin urmare, este esenţial să reglementăm aceste pieţe mai strict. Dacă Occidentul îşi doreşte să ţină pasul cu China în cursa pentru construirea IA, trebuie să primim cu braţele deschise o formă de capitalism mai blândă şi mai reglementată.

0011

Sfârşit

Această carte nu s-a terminat încă. Nu am vrut să identific doar provocările care ne vor aştepta în 2062, provocări ce vor fi introduse în mare parte de maşinării care pot să gândească mai bine decât noi, provocări pe care *Homo digitalis* va trebui să le rezolve. Acesta ar fi un sfârşit mult prea pesimist al unei poveşti despre ceva ce s-ar putea să fie una dintre cele mai importante creaţii ale noastre.

Cu toate acestea, ne apropiem de un moment critic în istoria omenirii. Ne confruntăm cu o multitudine de probleme pe lângă provocarea adusă de schimbările tehnologice – încălzirea climei globale, criza financiară globală, problema globală a refugiaţilor, şi acestea sunt doar câteva exemple. Într-adevăr, uneori pare că ne confruntăm numai cu probleme *globale.* Şi avem puţine cărţi pe care să le jucăm pentru a ne confrunta cu ele. Unul dintre puţinele lucruri pe care le putem face, dacă nepoţii noştri vor ajunge să ducă nişte vieţi mai bune decât noi, este exact ce au făcut bunicii noştri cu mulţi ani în urmă. Trebuie să acceptăm schimbările tehnologice. Dar trebuie să o facem cu precauţie, astfel încât vieţile noastre să fie îmbunătăţite.

O greşeală pe care o fac mulţi oameni este să considere că viitorul este fix, şi că pur şi simplu va trebui să ne adaptăm la el. Dar asta nu este adevărat. Viitorul este produsul deciziilor pe care le luăm în

prezent. Aşadar, putem alege viitorul pe care ni-l dorim.

Nu pretind că aş şti care sunt alegerile corecte. Nu eu trebuie să fiu cel care dă răspunsurile; sunt decizii pe care societatea ca întreg trebuie să le ia. În ciuda acestui lucru, voi încerca să identific câteva dintre alegerile pe care ar trebui să le facem.

Poate te va surprinde numărul de opţiuni pe care le-am identificat. E un lucru bun, dar şi rău. *Putem* să ne asigurăm că lumea anului 2062 va fi mai bună decât cea a anului 2018. În mod egal, sunt multe lucruri pe care *trebuie* să le facem ca să ne asigurăm că va fi aşa. Mai presus de toate, sunt convins că nu este posibil să lăsăm companiile de IT să continue să se reglementeze singure.

Companiile de tehnologie au făcut tentative disperate de a încerca să împiedice reglementarea sectorului lor; una este *Partnership on Artificial Intelligence*, lansat în luna septembrie a anului 2016[113]. Cu toate acestea, în ultimii ani au fost multe exemple prin care companiile de IT au demonstrat că nu putem avea încredere în ele că vor acţiona în interesul

[113] Parteneriatul Inteligenţei Artificiale pentru Beneficiul Oamenilor şi al Societăţii este o organizaţie înfiinţată de un consorţiu al actorilor din industria de IT, pentru a promova cele mai bune practici în domeniul IA. A fost fondat de Amazon, Facebook, Google, DeepMind, Microsoft şi IBM. Alte companii, inclusiv Apple, li s-au alăturat la scurt timp.

societăţii.

Nu e greu de văzut de ce avem nevoie de reglementări mai stricte în sectorul IT. Corporaţia este o instituţie de modă veche care nu se mai potriveşte cu binele public. Iar companiile de tehnologie în special reprezintă o provocare, luând în considerare efectele perturbatoare pe care noile tehnologii le au asupra vieţilor noastre.

Legi noi

Datele reprezintă una dintre zonele unde noi reglementări sunt necesare. Legile noi trebuie să limiteze colectarea şi folosirea datelor, de exemplu. Europa a arătat calea prin *Regulamentul General cu privire la Protecţia Datelor.* Legislaţie similară va fi necesară şi în alte ţări pentru a proteja intimitatea cetăţenilor. Dar nu trebuie să ne oprim doar la legi legate de protecţia datelor; va trebui să adoptăm şi altele care să controleze colectarea şi utilizarea datelor.

Am putea să ne gândim la legi care limitează drastic posesia de date. Poate că datele despre tine ar trebui să fie mereu ale tale. Poate vom decide că nimeni altcineva nu va mai avea permisiunea să deţină pulsul tău. ADN-ul ar trebui să fie şi el doar al tău, şi nu ceva ce poate fi cumpărat sau vândut de o corporaţie.

Majoritatea ţărilor deja au legi care previn vânzarea de părţi ale corpului – organe şi altele asemenea. S-ar putea să avem nevoie de legi asemănătoare care să împiedice vânzarea eului nostru digital, mai ales fără permisiunea unei persoane sau fără ca aceasta să beneficieze de ceva.

Poate că vom lua în considerare legi care să implice că datele pot să fie deţinute numai de cei care le creează. În cazul unei platforme precum Facebook, asta va însemna că utilizatorii vor deţine propriile postări, liste de prieteni, mesajele şi evenimentele. Utilizatorii vor putea atunci să îşi exporte datele pe alte reţele de socializare. Într-o astfel de lume, Facebook va avea nevoie să-şi facă utilizatorii fericiţi în continuare, altfel îi va pierde.

O altă măsură ar putea fi o lege a prescrierii folosirii de date de către o platformă sau o companie. Poate că limita ar trebui să fie 90 de zile? Sau un an? Dar la un moment dat, utilizatorii vor trebui să aibă dreptul să renegocieze termenii cu privire la felul în care sunt folosite datele lor. Şi va trebui ca utilizatorii să aibă dreptul ca datele lor să fie şterse.

Pe lângă legile legate de colectarea şi folosirea datelor, poate vom alege să reglementăm şi algoritmii care acţionează asupra acestor date. Aşa cum am văzut, algoritmii nu sunt neapărat nici corecţi, nici

transparenţi. Aşadar, poate că vom insista să fie. Şi poate că vom interzice complet ca algoritmii să fie folosiţi atunci când vine vorba de luarea unor decizii în legătură cu vieţile şi libertăţile oamenilor.

Pentru a încuraja inovaţia, platformele vor trebui să fie deschise la competiţie. Unele au devenit prea mari şi nimeni nu mai poate concura împotriva lor. Aşadar, trebuie să existe un ecosistem sănătos şi competitiv în cadrul acestor platforme. La fel cum multe state se asigură că firmele de utilităţi publice ne pun la dispoziţie ţevi şi cabluri la case pentru a crea o piaţă (oarecum artificială) pentru afacerile noastre, la fel va trebui să adoptăm legi pentru ca platformele digitale să permită competiţie în cadrul serviciilor lor.

În cele din urmă, avem nevoie de legi noi care să reglementeze platformele ca pe orice alt mediu unde poate fi publicat ceva. Platformele trebuie să fie răspunzătoare pentru conţinutul lor. Astfel vor fi forţate să fie mai active în privinţa problemelor pe care le-au tot evitat, cum ar fi ştirile false şi boţii.

Corporaţii noi

O altă decizie ar putea fi reformarea corporaţiilor. Faptul că multe tehnologii digitale sunt monopoluri naturale constituie o provocare fundamentală. Iar legile

existente nu sunt prea bune când vine vorba de monopolurile digitale. Încă de pe vremea lui Reagan, legea antitrust din Statele Unite a tolerat monopolurile atât timp cât nu au preţuri mai mari pentru consumatori. Legea antitrust a fost şi a rămas neputincioasă în privinţa reglementării companiilor ce oferă servicii gratis.

În prezent avem un număr mic de companii de tehnologie care domină spaţiul digital prin faptul că oferă servicii gratis. Aceste companii nu doar că domină comerţul de pe internet, căutările, reţelele de socializare, ci şi multe alte servicii, cum ar fi e-mail-ul, mesajele, filmuleţele şi reclamele digitale. Le-a fost permis chiar să-şi întărească monopolurile cumpărând companii rivale şi complementare. Google, spre exemplu, şi-a consolidat monopolul în domeniul video, cumpărând Youtube, în timp ce Facebook a cumpărat Instagram şi Whatsapp.

Aşa că de ce ne îngrijorăm? Cum poate un serviciu gratis să nu fie competitiv, chiar dacă este un monopol? Problema este că aceste servicii ascund adevăratul preţ faţă de consumatori. Când un serviciu pare că este „gratis”, preţul este inevitabil ascuns. Impactul serviciilor digitale asupra jurnalismului tipărit este un preţ. Dispozitivele noastre care dau dependenţă sunt un preţ. Pierderea intimităţii noastre are un preţ. Bulele

sociale au un preţ. Încetinirea inovaţiei de către companiile de IT mari care cumpără adversari mai mici şi mai agili are de asemenea un preţ.

Prin urmare, poate că va trebui să ne gândim cum să spargem companiile de tehnologie mari pentru a face ca piaţa digitală să fie din nou competitivă. De pildă, cu mult înainte de anul 2062, Alphabet va trebui să fie spart în bucăţi. Căutare. E-mail. Video. Sisteme de operare pentru dispozitive mobile. Niciunei companii nu ar trebui să îi fie permis să domine toate acestea. În mod ironic, prin înfiinţarea Alphabet, Google a făcut să fie şi mai uşor de desfăcut în bucăţi, făcând şi mai dificil de susţinut argumentele că trebuie să fie păstrat ca un tot unitar.

S-ar putea să trebuiască şi să îi împiedicăm pe giganţii de IT să cumpere sau să se unească cu competitorii lor. În 2012, Facebook a cumpărat Instagram pentru un miliard de dolari. Iar în anul 2014, a cumpărat Whatsapp, un competitor pentru serviciul său de mesaje, pentru 19 miliarde de dolari. Cu mult înainte de 2062 afaceri de genul acesta ar trebui să fie interzise. Instagram şi Whatsapp se descurcau foarte bine de unele singure, înainte ca Facebook să le cumpere. Şi probabil că Facebook a beneficiat cel mai mult, şi nu consumatorul.

Poate că revoluţia IA va face necesar să inventăm noi

tipuri de corporaţii pentru a ne asigura că vom avea parte cu toţii de beneficiile aduse de schimbările tehnologice. Astfel de corporaţii se vor armoniza mai bine cu valorile societăţii. Vor împărţi o parte mai mare din averea lor cu angajaţii şi cu clienţii lor. Şi vor putea să acţioneze pe termen lung, investind atât în angajaţii lor, cât şi în ţările în care operează.

În cele din urmă, va fi nevoie ca taxele să crească, astfel încât corporaţiile să contribuie la societate. Asta ar implica mai multă cooperare internaţională, astfel încât corporaţiile să nu poată asmuţi o ţară împotriva alteia. Ar putea însemna şi creşterea taxelor pe vânzări şi pe cifra de afaceri, întrucât acestea sunt şi mai greu de evitat, şi mai directe în ceea ce priveşte păstrarea averii acolo unde a fost obţinută.

Politică nouă

Reforma politică este o a treia decizie pe care va trebui să o luăm ca să facem lumea anului 2062 să fie aşa cum ne dorim. Scandalul Cambridge Analytica din anul 2018 a pus în lumină efectul coroziv pe care îl au datele în cantităţi masive asupra politicii. Pentru prevenirea câştigării competiţiilor politice de către cei care au algoritmi mai buni sau mai multe date, vom avea nevoie de legi care să limiteze folosirea datelor

pentru a influenţa opiniile politice ale oamenilor.

Ar trebui oare să prevenim transmiterea mesajelor politice unui public restrâns? Dacă ai o idee atractivă din punct de vedere politic, tot poţi să o difuzezi. Dar micro-ţintirea votanţilor de către algoritmi ar trebui să fie interzisă. Folosirea calculatoarelor ca „arme de convingere în masă" ar trebui să fie interzisă. Companiile care deţin reţele de socializare, de exemplu, ar trebui să nu mai aibă voie să vândă reclame politice pe baza niciunui criteriu specific, cu excepţia vârstei votantului şi a circumscripţiei. Nu vei mai putea să ai ca ţintă femei necăsătorite din generaţia Y sau bărbaţi albi pensionaţi cu mesaje politice care provoacă discordie. Dacă Facebook vrea să facă societatea să fie mai unită, poate că va trebui să nu mai vândă reclame în acest fel. Şi dacă refuză să o facă, va trebui să dăm legi ca să-i obligăm.

Poate că ne vom gândi şi la legi care să interzică complet folosirea reţelelor de socializare pentru obţinerea de voturi. Imaginaţi-vă o lume în care reclamele digitale ar putea să fie doar comerciale, şi nu politice. Un efect colateral pozitiv al unei astfel de schimbări ar fi că partidele politice vor fi obligate să-şi cheltuiască banii pe mijloacele mass-media tradiţionale, ajutând astfel jurnalismul de modă veche să supravieţuiască în era noastră digitală.

Poate că ne vom decide să interzicem şi boţii de pe internet. Nu prea avem motive să credem că boţii fac internetul un loc mai bun. Şi ar fi uşor să îi interzicem. Am putea pur şi simplu să dăm o lege prin care companiile de internet să trebuiască să verifice identitatea fiecărui utilizator, şi să fie amendate cu un anumit procent din venitul lor atunci când descoperă boţi care ne distrag atenţia. Vocile umane ar prelua din nou controlul pe internet.

Dar tot ne rămâne problema cu conţinutul fals: ştirile false, filmuleţele false şi înregistrările audio false. Tehnologia ne va veni în ajutor aici. Tehnologii precum blockchain pot ajuta la autentificarea datelor, în ciuda naturii descentralizate a internetului. Educaţia va ajuta şi ea. De exemplu, Italia învaţă acum copiii cum să recunoască ştirile false. Dar tot s-ar putea să avem nevoie de legislaţie. Dacă platformele sunt forţate să îşi asume responsabilitatea pentru conţinutul lor, atunci conţinutul fals va fi şters rapid.

Economii noi

Economia este un alt aspect al vieţii noastre care se va schimba până în anul 2062. Având în vedere impactul mare pe care IA îl va avea asupra locurilor de muncă, vor apărea câteva domenii economice cu totul noi. La

fel ca în cazul revoluţiei industriale, va fi esenţial ca angajaţii să fie protejaţi dacă vrem să ne asigurăm că 2062 va fi o lume mai bună pentru mulţi oameni, şi nu doar pentru câţiva.

Într-o perioadă în care avem atâta nevoie de ele, sindicatele din multe ţări par să fie lipsite de putere şi rămase în urmă. Este greu să ne imaginăm că se vor reinventa la timp pentru a putea face faţă revoluţiei IA. Dacă nu vor putea face asta, va trebui să apelăm la alte instituţii ca să ne asigurăm că vom beneficia cu toţii datorită schimbărilor tehnologice ce urmează să aibă loc.

Spre exemplu, am putea să dăm legi ca toate companiile să îi recalifice pe angajaţii care sunt concediaţi înainte să poată angaja alţii noi. Astfel de legi ar fi utile în prezent. La începutul anului 2018, banca australiană NAB a anunţat că a concediat 6000 de angajaţi din cauza digitalizării, şi a angajat 2000 care aveau abilităţi digitale. Un astfel de comportament crud şi pur capitalist ar trebui să fie interzis. NAB ar trebui în schimb să reinstruiască cel puţin 2000 din cei 6000 de angajaţii pe care aveau de gând să-i dea afară şi să le găsească locuri de muncă – poate că nu neapărat la NAB.

Am putea solicita companiilor să ofere personalului lor un nivel minim de educaţie şi de calificare

profesională. Ca să adaptăm o idee de la Google, am putea cere ca 20% din timpul angajaţilor să fie dedicat dezvoltării personale a acestora. Asta ar putea suna idealist – mai ales acel 20%. Dar cel mai bun atu al unei companii este adesea personalul său. Cum ar putea investirea în propria forţă de muncă să nu fie un plan bun pe termen lung pentru oricare afacere?

Asemenea idei nu îi ajută prea mult pe cei care lucrează pe cont propriu sau cu jumătate de normă într-o economie a precarităţii muncii. Deja vedem cum sunt abuzaţi aceşti oameni, aşa că va trebui cu siguranţă să introducem legi ca să îi protejăm. Asta se poate face şi cu frumosul, şi cu băţul. Am putea pretinde ca angajaţii part-time să primească multe dintre avantajele pe care le primesc angajaţii full-time, cum ar fi concediul medical şi cel parental. În mod egal, am putea face să fie mai atractiv din perspectiva taxelor pentru companii să angajeze oamenii, şi nu să le subcontracteze munca.

Am putea face ca balanţa puterii să fie în favoarea celor care îşi oferă munca. Abuzurile de genul contractelor cu zero ore lucrate ar putea fi interzise. Şi la fel cum o companie poate alege câte ore să ofere, la fel şi cei care primesc de lucru trebuie să poată decide câte ore acceptă, fără să le fie teamă de repercusiuni. Guvernul va trebui să intervină şi să ofere siguranţa pe

care o ofereau companiile cândva, astfel încât până şi cei care lucrează ocazional să poată beneficia de asigurare de sănătate, să îşi ia concediu medical, parental sau de îngrijire a unei persoane, şi să se poată recalifica la locul de muncă.

În cele din urmă, aceste pârghii punctuale ar putea fi inadecvate pentru a contracara efectele dispariţiei şi reducerii numărului locurilor de muncă tradiţionale care vor avea loc până în 2062. Aşadar, vom ajunge poate să trăim într-o economie în care venitul universal de bază a devenit ceva comun. Într-o variantă mai puţin radicală, poate că vom plăti pentru o bună parte din muncile care în ziua de azi rămân în general neplătite, cum ar fi îngrijirea celor bolnavi şi a celor bătrâni şi creşterea copiilor.

Poate că vom avea şi o săptămână de lucru mai scurtă. Din moment ce roboţii vor munci în locul nostru, anul 2062 ar putea fi o epocă a relaxării. Poate vom petrece cea mai mare parte a timpului alături de prieteni, în familii şi comunităţi. IA ne poate oferi aceste lucruri.

O societate nouă

Până în 2062 este posibil să trăim într-o societate mult mai blândă. Dacă îi vom aprecia pe cei care au grijă de

tineri, de bolnavi, de vârstnici şi de persoanele cu dizabilităţi la fel de mult sau poate chiar mai mult decât pe cei care au locuri de muncă tradiţionale, societatea noastră va fi una mult mai atentă cu nevoile tuturor. Astfel de meserii care presupun îngrijirea oamenilor nu sunt nici acum, şi probabil nu vor fi niciodată, pentru roboţi. Trebuie să începem să le preţuim mai mult.

Poate că vom vedea şi o înflorire a creativităţii. Aşa cum am sugerat anterior, ar putea fi vorba despre o A Doua Renaştere. Chiar dacă roboţii pot crea artă şi meşteşuguri, vom preţui mai mult obiectele şi experienţele create de oameni. Iar IA poate să ofere creşteri de productivitate care să ne permită să fim artişti şi artizani.

Presupunând că ne vom rezolva cumva problema prejudecăţilor şi părtinirii algoritmilor, până în anul 2062 vom avea şi o societate mult mai dreaptă. Automatele vor lua decizii fără a fi influenţate de prejudecăţile istorice şi culturale care ne-au distorsionat trecutul. Spre deosebire de oameni, aceste maşinării s-ar putea să ofere explicaţii raţionale pentru deciziile lor. Poate că vom implementa reglementări care să le ceară să facă acest lucru.

Societatea ar putea fi mai egalitară. Odată reglementată adecvat, tehnologia informaţiei în general şi IA în special pot fi minunate forţe egalizatoare. Mulţi

oameni vor scăpa de sărăcie. Şi mulţi se vor bucura de o viaţă confortabilă. Dar asta presupune că vom adopta măsuri astfel încât să avem cu toţii parte de beneficiile pe care le aduce IA, şi nu doar cei care deţin tehnologia.

În cele din urmă, societatea ar putea deveni şi mai paşnică. Presupunând că vom limita prin lege folosirea armelor letale autonome, IA ar putea chiar să salveze, şi nu să curme vieţi, curăţând terenurile minate, oferind ajutor umanitar, reducând numărul victimelor colaterale şi protejând soldaţii de pericole.

Dacă luăm deciziile corecte, IA aduce cu ea promisiunea unei vieţi mai bune – nu doar pentru câţiva, ci pentru toată lumea. Ne poate ajuta să trăim mai sănătos, să fim mai bogaţi şi poate chiar să ducem vieţi mai fericite.

Un viitor mai bun

În anul 2016 Barack Obama a spus că, „dacă ar trebui să alegi o perioadă din istoria omenirii în care să trăieşti, ai alege-o pe aceasta. Exact aici... exact acum”[114]. Ca să fiu sincer, de fapt a spus: „Exact aici, în America, exact acum”. Dar asta pentru că nu trăieşte într-un loc şi mai bun, cum ar fi Australia.

---

[114] Barack Obama, „Now Is the Greatest Time to Be Alive“, *Wired*, 10 decembrie 2016.

Există o doză de retorică în declaraţia lui Obama. Peste tot în lume speranţa de viaţă a crescut extrem de mult în secolul acesta. Cu toate astea, recent speranţa de viaţă din Statele Unite a început să scadă puţin, din cauza crizei opioidelor. Iar în timp ce mulţi oameni au scăpat de sărăcie, inegalitatea dinăuntrul multor ţări, în special din Statele Unite, a fost în creştere. Dar puse una peste alta, trăim aproape cu toţii vieţi mult mai bune acum decât strămoşii noştri în urmă cu o sută de ani.

Cum am ajuns să avem vieţi atât de bune? Cu ajutorul ştiinţei. Adesea auzi spunându-se: „Trăim în vremuri exponenţiale". Adepţii singularităţii îţi vor spune că tehnologiile exponenţiale ne schimbă lumea. Şi există o fărâmă de adevăr în asta. Dar poate că cea mai importantă schimbare exponenţială din timpul vieţilor noastre este una despre care nu s-a vorbit suficient de mult. Aceasta este progresul exponenţial al ştiinţei în sine. Belşugul pe care ni l-a oferit este motivul pentru care trăim vieţi mult mai bune acum decât în urmă cu un secol.

Pentru a parafraza cuvintele lui Isaac Newton, ştiinţa avansează exponenţial deoarece oamenii de ştiinţă stau pe umerii giganţilor. Putem beneficia de toate cunoştinţele ştiinţifice care au fost descoperite înainte de noi. Iar ştiinţa progresează exponenţial, de vreme ce

în prezent trăiesc mai mulţi oameni de ştiinţă decât câţi au existat în toate timpurile la un loc. Acestea sunt indicii că are loc un proces exponenţial.

Dar nu doar ştiinţa şi tehnologia pe care aceasta a făcut-o posibilă ne-au schimbat vieţile în ultimul secol. Am făcut şi schimbări semnificative în societate pentru a ne adapta la perturbările pe care progresul tehnologic ni le-a adus. Am introdus instituţii cum ar fi sindicatele, am creat dreptul muncii, educaţia universală şi conceptul de stat social astfel încât să avem cu toţii parte de prosperitatea adusă de evoluţia tehnologiei. Trebuie să ţinem minte acest lucru acum, când intrăm într-o nouă perioadă de transformări tehnologice profunde.

Inteligenţa artificială ne va schimba lumea dramatic. Lumea anului 2062 va fi foarte diferită de cea din prezent. Aşadar, trebuie să ne gândim cu atenţie la schimbările pe care ar trebui să le facem în societatea din prezent, dacă vrem să ne asigurăm că lumea anului 2062 va fi cea pe care ne-o dorim.

Să începem!

Mulţumiri

Aş dori să îi mulţumesc agentei mele, Margaret Gee, şi editurii mele, Black Inc.; munca lor a fost vitală

pentru ca această carte să ajungă în mâinile tale. Mulţumiri speciale lui Chris Feik, editorul meu; lui Diog Kagan şi lui Julian Welch, redactorii mei excelenţi şi înţelegători; lui Kim Ferguson pentru o copertă minunată; Christinei Taylor, lui Marian Blythe, lui Alison Alexanian şi lui Wilson da Silva pentru că au făcut publicitate fantastică; Nadiei Laurinci şi echipei sale de la Laurinci Speakers pentru că s-a ocupat de toate evenimentele mele de lansare; şi lui Sophy Williams pentru că a făcut o treabă extraordinară în privinţa drepturilor internaţionale.

Sunt mulţi alţi oameni cărora aş dori să le mulţumesc. Părinţilor mei, care m-au ajutat să mă lansez pe acest drum, acela de a visa la construirea inteligenţei artificiale.

Numeroşilor mei colegi cercetători de la UNSW Sydney, Data61 şi de oriunde altundeva, pentru că mi-au oferit mediul intelectual şi susţinerea care mi-au permis să scriu această carte, de la decan şi directorul şcolii până la colaboratorii şi studenţii mei.

Şi, mai presus de toate, vreau să mulţumesc întregii mele familii şi prietenilor, care îmi luminează fiecare zi şi care mi-au oferit timpul de a scrie *a doua* mea carte.

Bibliografie

Robert M. Bond, Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle & James H. Fowler, „A 61-million-person Experiment în Social Influence and Political Mobilization", *Nature*, vol. 489, 2012, pp. 295-298.

David Chalmers „Facing Up to the Problem of Consciousness", *Journal of Consciousness Studies*, vol. 2, no. 3, 1995, pp. 200-219.

David Chalmers, „The Singularity: A Philosophical Analysis", *Journal of Consciousness Studies*, vol. 17, nr. 9-10, 2010, pp. 7-65.

Era Dabla-Norris, Evridiki Tsounta, Kalpana Kochhar, Frantisek Ricka & Nujin Suphaphiphat, *Causes and Consequences of Income Inequality: A Global Perspective. Technical report*, Internaţional Monetary Fund, June 2015.

Amit Datta, Michael Carl Tschantz & Anupam Datta, „Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination", *Proceedings on Privacy Enhancing Technologies*, vol. 1, 2015, pp. 92-112.

Jared Diamond, *Collapse: How Societies Choose to Fail or Succeed*, Viking Press, New York, 2005.

Philippa Foot, *The Problem of Abortion and the Doctrine of the Double Effect în Virtues and Vices*, Basil Blackwell, Oxford, 1978. (Apărut pentru prima oară în *Oxford*

Review, nr. 5, 1967.)

Martin Ford, *The Lights the Tunnel: Automation, Accelerating Technology and the Economy of the Future,* Acculant Publishing, SUA, 2009.

Timnit Gebru, Jonathan Krause, Yilun Wang, Duyun Chen, Jia Deng, Erez Lieberman Aiden & Li Fei-Fei, „Using Deep Learning and Google Street View to Estimate the Demographic Makeup of Neighborhoods Across the United States", *Proceedings of the Naţional Academy of Sciences,* vol. 114, nr. 50, 2017, pp. 13108-13113.

Christof Henys, *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions,* Consiliul Organizaţiei Naţiunilor Unite pentru Drepturile Omului, 2013.

Martin Luther King Jr, *Where De We Go from Here: Chaos or Community?* Beacon Press, Boston, 1967.

Jakob Bæk Kristensen, Thomas Albrechtsen, Emil Dahl-Nielsen, Michael Jensen, Magnus Skovrind & Tobias Bornakke, „Parsimonious Data: How a Single Facebook Like Predicts Voting Behavior în Multiparty Systems", *PLoS One,* vol. 12, nr. 9, 2017.

James Manyika, Michael Chui, Mehdi Miremadi, Jacques Bughim, Katy George, Paul Willmott & Martin Dewhurst, *A Future that Works: Automation, Employment and Productivity,* Mekinsey Global Institute, 2017.

Ritesh Noothigattu, Neil Gaikwad, Edmund Awad, Sohad D'Souza, Iyad Rahwan, Pradeep Ravikumar & Ariel Procaccia, „A Voting-Based System for Ethical Decision Making", *Proceedings of 32nd AAAI Conference on Artificial Intelligence,* 2018.

Stephen Omohundro, „The Basic AI Drives", în Pei Wang, Ben Goertzel & Stan Franklin (eds), *Artificial General Intelligence 2008: Proceedings of the First AGI Conference, Frontiers în Artificial Intelligence and Applications 171,* pp. 483-492, IOS Press, Amsterdam, 2008.

Roger Penrose, *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics,* Oxford University Press, New York, 1989.

Steven Pinker, *The Better Angels of Our Nature,* Viking Books, New York, 2011.

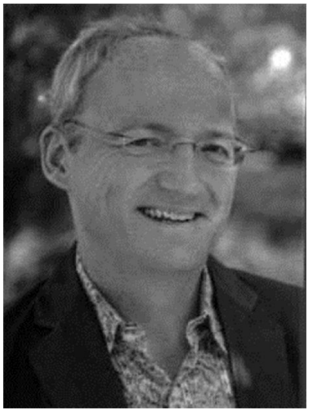
Richard Rhodes, *The Making of the Atomic Bomb,* New York, Simon & Schuster, 1986.

Brandon Schoettle & Michael Sivak, *A of Real-World Crashes Involving Self-Driving Vehicles,* The University of Michigan, Transportation Research Institute, Technical Report, 2015.

Alan Turing, „On Computable Numbers, with an Application to the Entscheidungsproblem", *Proceedings of the London Mathematical Society,* vol. 42, pp. 230-265, 1937.

Alan Turing, „Computing Machinery and Intelligence”, *Mind,* vol. 59, nr. 236, pp. 433-460, 1950.

Yilun Wang & Michal Kosinski, „Deep Neural Networks Are More Accurate than Humans at Detecting Sexual Orientation from Facial Images”, *Journal of Personality and Social Psychology,* vol. 114, nr. 2, pp. 246-257, 2018.



TOBY WALSH este unul dintre cei mai importanţi experţi din lume în domeniul inteligenţei artificiale şi profesor de IA la Universitatea din New South Wales, Australia. În 2015 a fost unul dintre oamenii din spatele unei scrisori deschise care solicita interzicerea armelor autonome, aşa-numiţii „roboţi ucigaşi”, semnată de peste 3000 de cercetători şi oameni de ştiinţă, antreprenori şi intelectuali de prim rang. Prima lui carte s-a numit în Australia It’s Alive! Artificial Intelligence from the Logic Piano to Killer Robots.

2062 este anul în care vom avea roboți la fel de inteligenți ca noi. Acest lucru este susținut de majoritatea experților în domeniile inteligenței artificiale și roboticii. Dar cum va arăta acest viitor? Cum se va desfășura viața pe această planetă?

Profesorul Toby Walsh analizează impactul pe care inteligența artificială îl va avea asupra muncii, războiului, politicii, economiei, vieții cotidiene și morții umane.

Pe baza unei înțelegeri profunde a tehnologiei și a implicațiilor acesteia, *2062: Lumea creată de inteligența artificială* descrie alegerile pe care trebuie să le facem astăzi pentru a ne asigura că viitorul va rămâne luminos.

ISBN 978-606-006-452-7
9 786060 064527
www.rao.ro
www.raobooks.com